



Universidad Politécnica de Valencia



Escuela Técnica Superior de Ingenieros  
Agrónomos

# **Estudios de estructura y de interacciones de LRRK2, una proteína implicada en la Enfermedad de Parkinson**

Marta Tormos Pérez

**Licenciatura en Biotecnología**

**Trabajo Final de Carrera**

Valencia 2009











Autor: Marta Tormos Pérez

DNI: 20454870G

**Título: Estudios de estructura y de interacciones de LRRK2, una proteína implicada en la Enfermedad de Parkinson**

Área o áreas de conocimiento a las que corresponde el trabajo: Biología Molecular

Titulación: Licenciatura en Biotecnología

Nombre y apellidos: Domingo Barettino Fraile

Departamento: Biotecnología

En calidad de:           director/a                                       codirector/a                                       **X** tutor/a

Autorizo la presentación del trabajo de fin de carrera cuyos datos figuran en el apartado anterior y certifico que se adecua plenamente a los requisitos formales, metodológicos y de contenido exigidos a un trabajo de fin de carrera, de acuerdo con la normativa aplicable en la ETSIA.

(Firma) Domingo Barettino Fraile

Valencia,

5 de

Noviembre

de 2009





UNIVERSIDAD  
POLITECNICA  
DE VALENCIA



ESCUELA TÉCNICA SUPERIOR  
DE INGENIEROS  
AGRÓNOMOS

#### Datos personales

Nombre y apellidos: Marta Tormos Pérez

#### Datos del trabajo de fin de carrera

Título del TFC: **Estudios de estructura y de interacciones de LRRK2, una proteína implicada en la Enfermedad de Parkinson**

Lugar de realización: Instituto de Biomedicina de Valencia

Fecha de lectura:

Titulación: Licenciatura Biotecnología

Director/a: Jordi Pérez Tur

Codirector/a: Fernando Cardona Serrate

Tutor/a: Domingo Barettino Fraile

#### **Resumen**

La enfermedad de Parkinson (EP) es una enfermedad neurodegenerativa para la que no se conoce cura y para la que se han propuesto como causas tanto factores genéticos como ambientales. El gen *LRRK2* (*Leucine-Rich Repeat Kinase 2*, *PARK8*) aparece mutado en una importante número de casos de la EP autosómica dominante y en algunos de los casos esporádicos de aparición tardía, por lo que el conocimiento de la función de la proteína LRRK2 proporciona una oportunidad para descifrar las rutas moleculares importantes en la neurodegeneración, particularmente en la que causa de la EP. LRRK2 es una proteína multidominio grande y compleja, que comprende varios dominios de interacción: repeticiones tipo armadillo, repeticiones tipo anquirina, repeticiones ricas en leucina (LRR) y un dominio WD40 C-terminal; dos dominios catalíticos: ROC-GTPasa y quinasa de la subfamilia de las treonina/serina quinasas; y un dominio de función desconocida: COR (C-terminal de ROC). Con el objetivo último de determinar las características estructurales de varios de estos dominios, en este trabajo se han ensayado diversas condiciones de expresión y extracción de los distintos dominios de LRRK2, clonados de manera independiente para obtener proteína soluble. También se ha realizado el repliegado *in vitro* del dominio COR, dado el interés que presenta la estructura y función de este dominio concreto. Además, se ha realizado un detallado análisis *in silico* de la estructura tridimensional de cada dominio de LRRK2 y del efecto que las mutaciones patogénicas y los polimorfismos descritos pueden tener en ésta. Este análisis nos ha permitido redefinir los límites de los diferentes dominios y se ha procedido a re-clonar los dominios COR y ROC-COR así definidos.

#### **Palabras clave**

Enfermedad de Parkinson, LRRK2, estructura, dominios.

## Resum

La malaltia de Parkinson (MP) és una malaltia neurodegenerativa per a la qual no es coneix cura i per a la que s'han proposat moltes causes incloent-hi factors genètics i també ambientals. El gen *LRRK2* (*Leucine-Rich Repeat Kinase 2*, PARK8) apareix mutat en un important nombre de casos de la MP autosòmica dominant i en alguns casos esporàdics d'aparició tardana, pel que el coneixement de la funció de la proteïna LRRK2 proporciona una oportunitat per a desxifrar les rutes moleculars importants en la neurodegeneració, particularment en la causa de la MP. LRRK2 és una proteïna multidomini gran i complexa, que conté diversos dominis d'interacció: repeticions tipus armadillo, repeticions tipus anquirina, repeticions riques en leucina (LRR) i un domini WD40 C-terminal; dos dominis catalítics: ROC-GTPasa i quinasa de la subfamília de les treonina/serina quinases; i un domini de funció desconeguda: COR (C-terminal de ROC). Amb l'objectiu final de determinar les característiques estructurals de diversos d'estos dominis, en aquest treball s'han assajat diferents condicions d'expressió i extracció dels diferents dominis de LRRK2, clonats de forma independent per obtenir proteïna soluble. També s'ha realitzat el replegat *in vitro* del domini COR, degut a l'interés que presenta l'estructura i funció d'aquest domini en concret. A més, s'ha realitzat un detallat anàlisi *in silico* de l'estructura de cada domini de LRRK2 i els efectes que les mutacions patogèniques i els polimorfismes descrits pogueren tindre sobre aquesta. Aquest anàlisi ens ha permès redefinir els límits dels diferents dominis, i s'ha procedit a re-clonar els dominis COR i ROC-COR aixina definits.

## Paraules clau

Malaltia de Parkinson, LRRK2, estructura, dominis.

## Abstract

Parkinson's disease (PD) is a neurodegenerative disease for which there is not known cure, and for which have been proposed as causes genetic and environmental factors. The *LRRK2* gene (*Leucine-Rich Repeat Kinase 2*, PARK8) is mutated in a significant number of cases of autosomal dominant PD and in some sporadic cases of the late-onset type, so that knowledge of function of the protein provides an opportunity for deciphering molecular pathways important in neurodegeneration, particularly in which causes PD. LRRK2 is a large and complex multidomain protein that comprises several domains: armadillo-type repeats, ankyrin-type repeats, leucine-rich repeats (LRR) and a C-terminal WD40 domain, two catalytic domains: ROC-GTPase and a kinase of the threonine/serine kinases subfamily, and a domain of unknown function: COR (C-terminal of ROC). With the ultimate goal of determining the structural characteristics of several of those domains, in this work we have tested various conditions of expression and extraction of different domains of LRRK2, independently cloned, to obtain soluble protein. I have also been performing *in vitro* refolding of the COR domain, due to the interest on the structure and function of this particular domain. Furthermore, we have made a detailed *in silico* analysis of the three-dimensional structure of each domain of LRRK2 and the effect that pathogenic mutations and described polymorphisms may have on it. This latter analysis has allowed us to redefine the limits of the different domains and prompted the re-cloning ROC-COR and COR re-defined domains.

## Key words

Parkinson's disease, LRRK2, structure, domains.





# ÍNDICE

<b>1. INTRODUCCIÓN.....</b>	<b>1</b>
<b>1.1. Enfermedad de Parkinson .....</b>	<b>1</b>
1.1.1. Patología .....	1
1.1.2. Prevalencia.....	2
1.1.3. Etiología de la EP.....	2
1.1.4. Genes implicados en la EP .....	3
<b>1.2. LRRK2/dardarina .....</b>	<b>4</b>
1.2.1. Localización .....	4
1.2.2. Dominios y estructura.....	5
1.2.2.1. Estudios estructurales en LRRK2 .....	7
1.2.3. Genes homólogos y estudios en sistemas modelo.....	10
1.2.4. Función e interacciones con otras proteínas.....	11
1.2.5. Mutaciones de LRRK2 asociadas a la EP .....	13
1.2.5.1. Efectos de las mutaciones de LRRK2 a nivel celular .....	15
<b>1.3. JUSTIFICACIÓN Y OBJETIVOS.....</b>	<b>16</b>
<b>2. MATERIALES Y MÉTODOS .....</b>	<b>19</b>
<b>2.1. Material biológico .....</b>	<b>19</b>
2.1.1. Cepas bacterianas .....	19
2.1.2. Plásmidos utilizados.....	19
2.1.3. Medios y condiciones de cultivo.....	20
<b>2.2. Métodos de manipulación y análisis de proteínas .....</b>	<b>21</b>
2.2.1. Geles de acrilamida y tinción de Coomassie.....	21
2.2.2. Cuantificación de proteínas .....	21
2.2.3. Expresión y obtención de proteínas .....	21
2.2.4. Obtención y purificación de cuerpos de inclusión.....	21
2.2.5. Cromatografía de afinidad con resina de Ni .....	22
2.2.6. Replegado proteico <i>in vitro</i> .....	23
2.2.7. Cromatografía de exclusión molecular .....	24
2.2.8. Prueba de estabilidad proteica .....	25
2.2.9. Obtención de extracto de cerebro de rata .....	26
2.2.10. Ensayos de <i>pull-down</i> .....	26
2.2.11. Concentración de proteínas por precipitación con acetona .....	27
2.2.12. Análisis <i>in silico</i> de la estructura de los dominios de LRRK2.....	27
<b>2.3. Métodos de manipulación y análisis de ácidos nucleicos .....</b>	<b>28</b>
2.3.1. Electroforesis de DNA en geles de agarosa .....	28
2.3.2. Amplificación y preparación del inserto .....	28
2.3.3. Producción de bacterias <i>E. coli</i> DH5α electrocompetentes .....	29
2.3.4. Ligación y transformación.....	29
2.3.5. Comprobación de la ligación, extracción del DNA plasmídico y obtención de los clones.....	30
<b>2.4. Herramientas bioinformáticas .....</b>	<b>30</b>
<b>3. RESULTADOS Y DISCUSIÓN .....</b>	<b>31</b>
<b>3.1. Trabajo con dominios proteicos de LRRK2 .....</b>	<b>31</b>
3.1.1. Determinación de las características físico-químicas .....	31
3.1.2. Expresión y pruebas de solubilidad .....	31

3.1.3. Replegado <i>in vitro</i> .....	34
3.1.4. Prueba de estabilidad proteica .....	38
3.1.5. Ensayo de <i>pull-down</i> .....	39
<b>3.2. Análisis <i>in silico</i> de la estructura terciaria de los dominios de LRRK2 .....</b>	<b>40</b>
3.2.1. Análisis del dominio armadillo .....	42
3.2.2. Análisis del dominio anquirina .....	43
3.2.3. Análisis del dominio LRR .....	45
3.2.4. Análisis del dominio ROC .....	47
3.2.5. Análisis del dominio COR .....	50
3.2.6. Análisis del dominio quinasa .....	52
3.2.7. Análisis estructural del dominio WD40 .....	54
<b>3.3. Clonación de dominios de LRRK2 .....</b>	<b>59</b>
<b>4. CONCLUSIONES .....</b>	<b>61</b>
<b>5. BIBLIOGRAFÍA .....</b>	<b>63</b>
<b>ANEXOS .....</b>	<b>71</b>
Anexo 1_ Cuantificación proteica .....	73
Anexo 2_ Calibrado de la cromatografía de exclusión molecular .....	75
Anexo 3_ Tampones utilizados en la prueba de estabilidad proteica .....	77
Anexo 4_ Análisis <i>in silico</i> : definición de dominios .....	78
Anexo 5_ Análisis <i>in silico</i> : estructura secundaria (JPred3) .....	80
Anexo 6_ Análisis <i>in silico</i> : estructura terciaria .....	83

---







# 1. INTRODUCCIÓN

---

## 1.1. Enfermedad de Parkinson

La enfermedad de Parkinson (EP), descrita por primera vez en 1817 por James Parkinson, es una afección crónica de progresión continua para la que no se conoce cura. Se caracteriza por provocar temblores en reposo, bradiquinesia, rigidez e inestabilidad postural, y está causada principalmente por la pérdida progresiva de las neuronas dopaminérgicas de la *substantia nigra* del Sistema Nervioso Central. Además de la afectación de la vía dopaminérgica, encargada del control del movimiento muscular, esta patología provoca el desequilibrio de otros neurotransmisores como GABA, acetilcolina o serotonina, lo que se vincula con los síntomas no motores que presenta esta enfermedad (depresión, lentificación de los procesos cognitivos, fatiga, ansiedad, trastornos del sueño, etc.) y que en muchos casos son los que en mayor medida reducen la calidad de vida del paciente.

La EP se caracteriza por presentar una buena respuesta inicial al tratamiento con levodopa, precursor de la dopamina, que actualmente es la droga más efectiva para el alivio de la sintomatología de la enfermedad. Sin embargo, conforme progresa la patología y el desarrollo de los síntomas, la eficacia de este fármaco disminuye (Lewitt, 2008), provocando además la aparición de efectos secundarios asociados al exceso de dopamina en el cerebro, principalmente movimientos anormales e involuntarios (discinesia).

### 1.1.1. Patología

A nivel celular, la patología típica en la EP y otros parkinsonismos (aunque no aparece en todos los casos) es la formación de unas inclusiones citoplasmáticas anormales compuestas por agregados proteicos denominadas cuerpos de Lewy. Aún se desconoce la causa de la aparición de estas inclusiones, pero es sabido que su presencia desencadena la degeneración neuronal. Los cuerpos de Lewy se encuentran fundamentalmente en las neuronas dopaminérgicas y, aunque se desconoce su composición exacta, se han identificado más de 70 tipos de moléculas distintas en

estos agregados (Wakabayashi *et al.*, 2007). También es común en esta enfermedad la presencia prolongaciones fibrilares en las neuronas supervivientes (neuritas de Lewy).

Sin embargo, los mecanismos moleculares concretos que conducen a la neurodegeneración en pacientes de Parkinson están todavía por determinar. Se ha sugerido que el estrés oxidativo, la disfunción mitocondrial, los defectos en el sistema de ubiquitinación y degradación de proteínas, así como la formación de agregados celulares y la apoptosis; pueden ser procesos relevantes en la patogénesis de la enfermedad (Yang *et al.*, 2009).

### **1.1.2. Prevalencia**

La EP es la segunda enfermedad neurodegenerativa más común después de la enfermedad de Alzheimer, afectando al 1-2 % de la población de alrededor de 65 años y aumentando su incidencia con la edad. De hecho, la prevalencia de la enfermedad en países industrializados es del 0,3 % de la población general, y llega a alcanzar el 3,5 % de la población entre 65 y 89 años (De Rijk *et al.*, 2000); apareciendo muy pocos casos antes de los 50 años (De Lau y Breteler, 2006).

### **1.1.3. Etiología de la EP**

Aunque ha sido descrito el origen genético de la EP en algunos casos, también los factores ambientales como infecciones virales, neurotoxinas o algunos pesticidas (por ejemplo rotenona o paracuato), pueden actuar como factores de riesgo o incluso provocar directamente la enfermedad (Khandhar y Marks, 2007; Jang *et al.*, 2008). Debido a estos hechos en un principio se postuló que la EP se debía principalmente a factores ambientales. Sin embargo, la identificación de múltiples afectados en cuatro generaciones de dos familias procedentes de una pequeña aldea italiana (Contursi), y el posterior estudio de varias familias en las que la patología tenía una herencia mendeliana autosómica dominante, contribuyó a la consideración de la etiología genética de la enfermedad (Golbe *et al.*, 1990; Waters y Miller, 1994). Actualmente, se conoce que aproximadamente el 5-10 % de los casos de EP presentan herencia de tipo mendeliano, indicando claramente un origen genético. No obstante, la mayor parte de los casos son esporádicos, combinándose factores de riesgo genéticos y/o ambientales.

#### 1.1.4. Genes implicados en la EP

Gracias a numerosos estudios que asocian la EP con factores genéticos, se han caracterizado 14 *loci* relacionados con la aparición de diferentes variantes de la enfermedad (Biskup *et al.*, 2008; Yang *et al.*, 2009; Kumari y Tan, 2009). Aunque mutaciones en algunos de estos genes explican la aparición de la enfermedad en un bajo porcentaje de enfermos, han suscitado gran interés debido a que su estudio podría ayudar a la comprensión de las bases moleculares de esta patología.

En la Tabla 1 se muestran los *loci* implicados en la aparición de EP o parkinsonismos, aunque cabe destacar que existen casos familiares en los que no se han encontrado mutaciones en ninguno de los *loci* descritos. Además, de los 14 *loci* descritos no todos son de herencia mendeliana, si no que algunos presentan una herencia compleja/indeterminada.

Tabla 1.1: *Loci* relacionados con la EP.

<i>Locus</i>	Cromosoma	Gen	Herencia	Descripción	Edad de inicio
<b>PARK1</b>	4q21	<i>SNCA</i> /α-sinucleína	AD	Función sináptica	40s
<b>PARK2</b>	6q25	<i>PRKN</i> /parkina	AR	E3 ubiquitina ligasa	20-40
<b>PARK3</b>	2p13	¿?	AD	¿?	60s
<b>PARK5</b>	4p14	<i>UCHL1</i>	AD	Degradación proteasomal	~50
<b>PARK6</b>	1p35-p37	<i>PINK-1</i>	AR	Quinasa mitocondrial	30-40
<b>PARK7</b>	1p38	<i>DJ-1</i>	AR	Chaperona. Respuesta a estrés	30-40
<b>PARK8</b>	12q12	<i>LRRK2</i> /dardarina	AD	Quinasa/GTPasa	~60
<b>PARK9</b>	1p36	<i>ATP13A2</i>	AR	Bomba iónica lisosomal	20-40
<b>PARK10</b>	1p32	¿?	AD (¿?)	¿?	50-60
<b>PARK11</b>	2q34	<i>GIGYF2</i>	AD (¿?)	¿?	Tardío
<b>PARK12</b>	Xq31	¿?	Ligada al X	¿?	Tardío
<b>PARK13</b>	2p12	<i>OmiHtrA2</i>	¿?	Proteasa mitocondrial	Tardío
<b>PARK14</b>	22q13.1	<i>PLA2G6</i>	AR	Fosfolipasa A2 calcio independiente	20-40

<b>Locus</b>	<b>Cromosoma</b>	<b>Gen</b>	<b>Herencia</b>	<b>Descripción</b>	<b>Edad de inicio</b>
<b>PARK15</b>	22q12-q13	FBXO7	AR	¿?	10-20

AD: autosómica dominante. AR: autosómica recesiva.\* El *locus* PARK4 fue descrito inicialmente en una única familia, pero estudios posteriores demostraron que esta familia presentaba la enfermedad debido a la existencia de una triplicación del gen *SNCA*; por lo que este *locus* no se indica en la tabla.

## **1.2. LRRK2/dardarina**

En el año 2004, se identificaron por primera vez mutaciones en el gen *LRRK2* (*leucine-rich repeat kinase2*) como responsables de formas familiares autosómicas dominantes y en formas esporádicas de la EP (Paisan-Ruiz *et al.*, 2004). El gen *LRRK2* se localiza en el cromosoma 12 y está compuesto por 51 exones. Su producto génico es una proteína multidominio de 2527 aminoácidos y aproximadamente 280KDa, denominada LRRK2 o dardarina (del vasco “dardara” que significa temblor).

Hoy en día se considera que algunas mutaciones concretas en dardarina, como p.R1441G, p.G2019S y otras, son una de las principales causas de la enfermedad, siendo responsable de aproximadamente el 2-4 % de los casos en poblaciones caucásicas (Huang *et al.*, 2004). Dado esta elevada prevalencia el interés por su estudio ha ido en aumento desde su descubrimiento, siendo actualmente una de las principales líneas de investigación en cuanto a las causas moleculares de la EP.

### **1.2.1. Localización**

Se ha visto la expresión del gen *LRRK2* en una amplia variedad de tejidos (hígado, pulmones, riñones, corazón y cerebro), aunque a bajos niveles. En el cerebro *LRRK2* se expresa en distintas poblaciones neuronales, entre las que se incluyen las de la corteza cerebral, el *caudatus-putamen* y el hipocampo, así como en las neuronas dopaminérgicas de la *substantia nigra*, es decir, se encuentra presente tanto en zonas cerebrales relacionadas con la EP como en zonas no relacionadas (Simón-Sánchez *et al.*, 2006).

A nivel celular, LRRK2 es fundamentalmente citoplásmica y se encuentra asociada a las membranas de orgánulos celulares como la mitocondria, el retículo endoplasmático, los endosomas y las vesículas sinápticas, y también a la membrana

plasmática. Además se ha detectado su presencia en los sinaptosomas de las neuritas (Hatano *et al.*, 2007). Cabe destacar que LRRK2 presenta una distribución celular similar a la de otras proteínas relacionadas con la EP, como son la  $\alpha$ -sinucleína y la parkina (West *et al.*, 2005).

### 1.2.2. Dominios y estructura

El análisis de la secuencia de aminoácidos de LRRK2 indica que está formada por varios dominios independientes, algunos de los cuales son típicos de interacción entre proteínas: armadillo (ARM), anquirina (ANK), dominios ricos en leucina (LRR) y dominios WD40; y dos son dominios catalíticos: GTPasa (ROC) y quinasa (KIN). Además posee un dominio de función desconocida (COR, “C-terminal of ROC”), situado entre los dominios GTPasa y quinasa (Santpere y Ferrer, 2009).

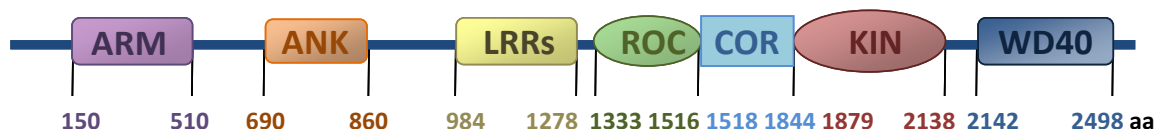


Figura 1.1. Representación esquemática de la estructura de dominios de LRRK2.

La región N-terminal de LRRK2 contiene siete repeticiones de tipo anquirina, que consisten cada una de ellas en dos hélices antiparalelas unidas por una lámina- $\beta$  o por un lazo, cuya unión forma el dominio ANK. Este tipo de dominio se ha encontrado en diversas proteínas eucariotas y procariotas, incluyendo proteínas citoesqueléticas, factores de transcripción, proteínas de señalización y reguladores del ciclo celular (Mata *et al.*, 2006). También se han identificado trece repeticiones ricas en leucina (LRR) predichas como una lámina- $\beta$  seguida de una hélice- $\alpha$ , este dominio que participa en interacciones con otras proteínas a través de la extensa superficie que presenta accesible al solvente (Mata *et al.*, 2006). La presencia o no de un dominio de interacción de repeticiones tipo armadillo en la región N-terminal ha sido controvertida. Sin embargo, estudios recientes parecen concluir que LRRK2 contiene repeticiones tipo armadillo (ARM-like) formadas por tres hélices- $\alpha$  que interaccionan formando una superhélice (Marín, 2008). También se han predicho repeticiones en C-terminal que formarían un dominio WD40. Este dominio ha sido identificado en reguladores transcripcionales, subunidades de fosfatasas, proteínas ensambladores del citoesqueleto y en proteínas involucradas en la formación y tráfico de vesículas. A

pesar de la diversidad funcional de las proteínas que contienen las repeticiones de WD40, la estructura tridimensional del dominio está muy conservada basándose en repeticiones de cuatro láminas- $\beta$  antiparalelas conectadas mediante lazos ( *$\beta$ -propeller*) (Mata *et al.*, 2006). Es interesante destacar que los dominios LRR y WD40 presentan una elevada carga neta positiva (+20 y +21, respectivamente, a pH fisiológico) debido principalmente a la presencia de residuos de arginina. En el dominio LRR las cargas positivas están situadas en la superficie, por lo que es posible que LRR y/o WD40 unan proteínas cargadas negativamente, fosfolípidos o ácidos nucleicos (Mata *et al.*, 2006).

Por otra parte, *LRRK2* codifica dos actividades enzimáticas: quinasa y GTPasa. El dominio quinasa, presenta un pequeño lazo N-terminal y un gran lazo C-terminal conectados por una región en forma de hendidura donde se une el  $Mg^{+2}$ -ATP y el sustrato. La activación de la quinasa ocurre en una secuencia de 25-30 aminoácidos del lazo C-terminal que está flanqueada por motivos conservados DF/YG y APE. Se requiere la fosforilación del segmento de activación para que la quinasa sea activa, de forma que el lazo C-terminal adquiere una conformación que permite el acceso al sustrato para que tenga lugar la catálisis (Mata *et al.*, 2006). Basándose en estudios de similitud de secuencia, se ha descrito que el dominio quinasa de *LRRK2* pertenece a la subfamilia TKL de las serina/treonina quinasas humanas. El hecho de que muchas quinasas de esta familia son activados por GTPasas pequeñas, concuerda con el hecho de que la actividad quinasa de *LRRK2* esté regulada intramolecularmente por la unión de GTP al dominio ROC-GTPasa (Ito *et al.*, 2007; Guo *et al.*, 2007). También se ha observado que la actividad quinasa fosforila el dominio GTPasa (Greggio *et al.*, 2009) y se ha postulado que existe una cascada de activación (transducción de señales) dentro de la proteína (van Egmond *et al.*, 2008; Marin *et al.*, 2008). El dominio GTPasa de *LRRK2* pertenece a la familia ROCO, en la cual la GTPasa predicha se encuentra siempre en tándem con el dominio COR, cuya función se desconoce. Este módulo ROC-COR se conserva a lo largo de la evolución, tanto en procariotas como en eucariotas, sugiriendo la interdependencia funcional de los dos dominios. El dominio ROC-GTPasa de *LRRK2* parece cercano a la familia RAB de GTPasas, que están implicadas en el tráfico y transporte vesicular (Mata *et al.*, 2006). Además, se ha visto que el dominio

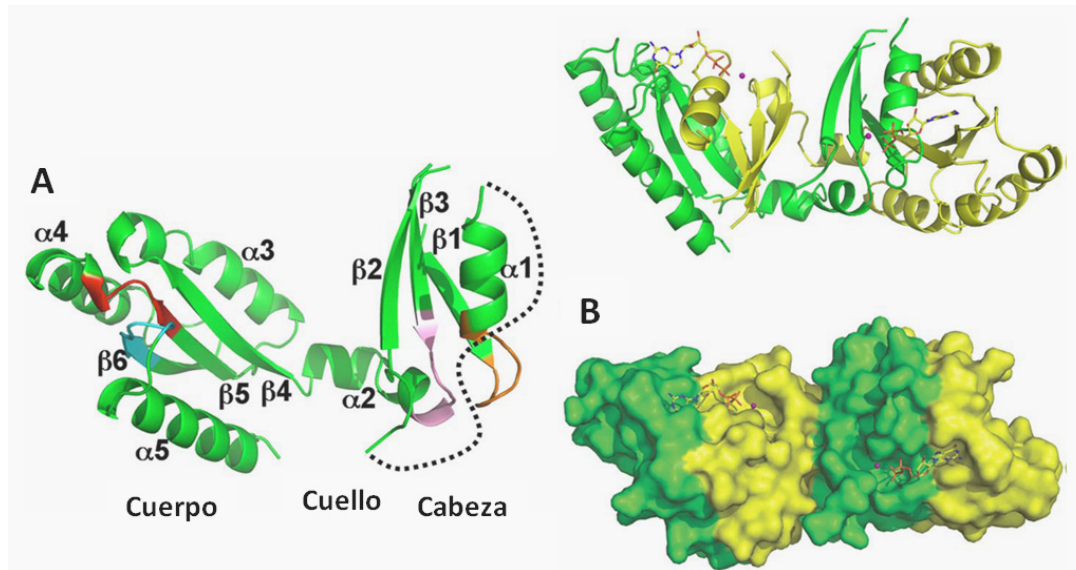


quinasa catalítico activo mínimo incluye (además del dominio quinasa) a la GTPasa y al dominio COR (Jaleel *et al.*, 2007).

Probablemente LRRK2 se encuentra predominantemente formando un dímero *in vivo*. La dimerización es un fenómeno bien establecido entre las quinasas, y en este caso tiene lugar a través de varias regiones de LRRK2, aunque qué partes están involucradas en el fenómeno es un tema en controversia (Greggio *et al.*, 2008). Algunos autores postulan que el dominio ROCO es crítico, aunque no exclusivo, para la dimerización (Klein *et al.*, 2009). Otras proteínas de la familia de ROCO también forman dímeros mediante interacciones COR-COR, aunque estas interacciones no se han observado en LRRK2, en la que sí hay evidencias de interacciones ROC-ROC obtenidas a partir de la estructura tridimensional (Deng *et al.*, 2008). Recientemente se ha descrito que, en estado inactivo, LRRK2 se encuentra en forma de oligómeros de gran tamaño estabilizados por chaperonas, incrementándose la formación de estructuras diméricas al activar los dominios GTPasa y quinasa de la proteína (Sen *et al.*, 2009).

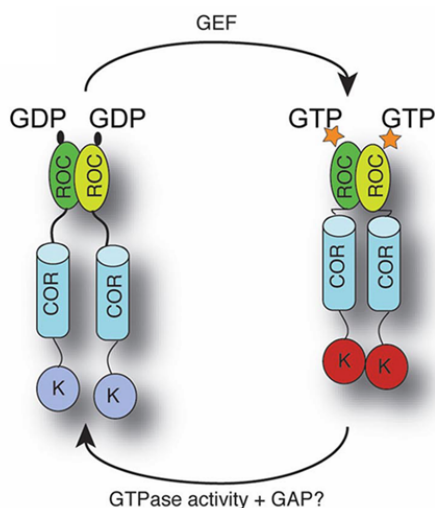
#### 1.2.2.1. Estudios estructurales en LRRK2

La estructura del dominio ROC de LRRK2 humana ha sido resuelta mediante cristalografía de rayos X en complejo con GDP-Mg<sup>2+</sup> (código PDB: 2zej; Deng *et al.*, 2008). Según estos resultados el dominio ROC forma un homodímero donde cada monómero contiene cinco hélices- $\alpha$  y seis láminas- $\beta$  unidas mediante lazos, y muestra tres subdominios: “cabeza”, “cuello” y “cuerpo”. La “cabeza” está formada por  $\beta 1$ ,  $\alpha 1$ ,  $\beta 2$  y  $\beta 3$  y los lazos entre  $\alpha 2$  y  $\beta 2$  y entre  $\beta 2$  y  $\beta 3$  desestructurados. El “cuello” está formado principalmente por una hélice curvada,  $\alpha 2$ , mientras que el “cuerpo” está constituido por  $\beta 4$ ,  $\alpha 3$ ,  $\beta 5$ ,  $\alpha 4$ ,  $\beta 6$  y  $\alpha 5$ . La “cabeza” y la primera parte del “cuello” de uno de los monómeros interacciona con el “cuerpo” del otro formando dos unidades compactas (ROC1 y ROC2).



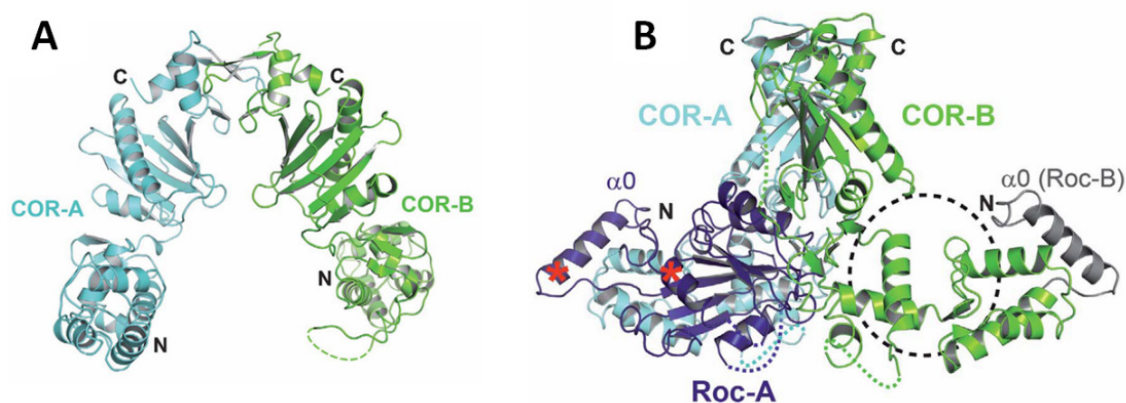
**Figura 1.2. Representación de la estructura dimérica del dominio ROC-GTPasa de LRRK2.** A) Representación de un dominio ROC en estado de monómero indicando sus tres subdominios “cabeza”, “cuello” y “cuerpo”, y las distintas estructuras secundarias. B) Representaciones del dímero ROC-ROC unido a GDP-Mg<sup>2+</sup> (modificado de Deng *et al.*, 2008).

El modelo propuesto por Deng *et al.* (2008) sugiere que los residuos R1441 e I1371, asociados a la EP, se localizan en la interfase de los dos monómeros y desestabilizan el dímero de ROC. Los datos de la estructura sugieren que la pérdida de las fuerzas estabilizadoras del dímero está relacionada con la disminución de la actividad GTPasa, y que durante el ciclo de unión GTP/GDP los cambios conformacionales del dominio ROC pueden ser transducidos a la quinasa mediante el dominio COR dirigiendo la dimerización del dominio quinasa y por lo tanto su autofosforilación y activación.



**Figura 1.3. Modelo de la regulación de la actividad quinasa a través del dominio ROC.** La unión de GTP activa a ROC induciendo la dimerización de COR, lo que produce la asociación del dominio quinasa y su autofosforilación y activación (color rojo). Por otro lado, la hidrólisis de GTP a GDP provoca un cambio conformacional de ROC que rompe la asociación dimérica de COR y el dominio quinasa, inactivando a la quinasa (color azul). Se desconocen la proteína activadora de GTPasa (GAP) y el factor intercambiador de guanina (GEF) implicados (Deng *et al.*, 2008).

Por otra parte, también se ha resuelto mediante difracción de rayos X la estructura del ROC-COR de *Chlorobium tepidum* (código PDB: 3dpu; Gotthardt *et al.*, 2008), proteína procariota constituida por los dominios LRR, ROC y COR. Los resultados muestran que el dominio COR de *C. tepidum* presenta dos subdominios. El N-terminal consiste principalmente en un dominio hélice- $\alpha$  con tres pequeñas láminas- $\beta$  antiparalelas, mientras el C-terminal contiene siete láminas- $\beta$  antiparalelas flanqueadas por cuatro hélices y un motivo  $\beta$ -hairpin (dos láminas- $\beta$  antiparalelas unidas mediante un lazo de 2-5 aas). La proteína dimeriza a través del subdominio C-terminal de COR, en ambos monómeros.



**Figura 1.4. Representación de la estructura dimérica de los dominios ROC-COR de *C. tepidum*.** A) Representación de la estructura del homodímero de COR interaccionando por la subunidad C-terminal en ambos monómeros. B) Representación del homodímero ROC-COR (modificado de Gotthardt *et al.*, 2008).

La interfase ROC-COR es el área más conservada de ambos dominios, lo que concuerda con la conservación del tándem ROC-COR en distintas proteínas a lo largo de la evolución. Mutaciones análogas a las de LRRK2 humana en ROC-COR de *C. tepidum* se encuentran próximas a dicha interfase y disminuyen el ratio de actividad GTPasa, probablemente debido a la modificación de la interacción entre estos dominios (Gotthardt *et al.*, 2008).

Cabe destacar que exceptuando el caso del dominio ROC, para el cual sí ha sido resuelta la estructura tridimensional, y para el dominio ROC-COR del procariota *C. tepidum*, los análisis estructurales de la dardarina humana realizados hasta la fecha

están basados en predicciones (Mata *et al.*, 2006), de modo que ni siquiera la definición de dominios es necesariamente exacta.

### **1.2.3. Genes homólogos y estudios en sistemas modelo**

El gen *LRRK1*, parálogo a *LRRK2*, codifica una proteína que presenta una disposición de dominios, patrón de expresión y tamaño similares a *LRRK2*, aunque no se ha asociado con la EP (Taylor *et al.*, 2007). También en la regulación catalítica existen similitudes, ya que se sabe que en *LRRK1* la actividad quinasa está regulada por GTP (Korr *et al.*, 2005). Las principales diferencias entre las proteínas se dan en el extremo N-terminal, de mayor tamaño en *LRRK2* (Greggio y Cookson, 2009).

Los genes *LRRK1* y *LRRK2* se encuentran conservados en vertebrados y tienen un ortólogo tanto en *Drosophila melanogaster* (*LRRK*) como en *Caenorhabditis elegans* (*lrk-1*) (Mata *et al.*, 2006). El trabajo con los homólogos en estos sistemas modelo ha permitido conocer características de la proteína humana, así como respaldar estudios realizados sobre la misma. Se han generado y caracterizado transgénicos de pérdida de función de *LRRK2* en *D. melanogaster*. Estos transgénicos tienen alterada su actividad motora y presentan degeneración de sus neuronas dopaminérgicas. No obstante, la sobreexpresión de la proteína silvestre o de los mutantes implicados de la EP no muestra anomalías significativas. En conjunto, estos resultados muestran que *LRRK* es crítico para la integridad de las neuronas dopaminérgicas y la actividad locomotora intacta en *D. melanogaster* (Lee *et al.*, 2007). Además se han generado y caracterizado mutantes de *C. elegans* de pérdida de función de *lrk-1*. Estos animales tienen afectada la distribución proteica de las vesículas sinápticas, lo que podría resultar en la degeneración neuronal debida al estrés metabólico y a alteraciones en la neurotransmisión (Melrose, 2008; Yue, 2009). Otros estudios describen que la ausencia de *lrk-1* en *C. elegans* afecta a la función mitocondrial, y aumenta la sensibilidad al estrés oxidativo (Sämann *et al.*, 2009). También se ha visto que la expresión de *LRRK2* silvestre en *C. elegans* aumenta la supervivencia del nematodo tras ser expuesto a rotenona o paracuato (compuestos que causan disfunción mitocondrial y estrés oxidativo), mientras que la expresión de formas mutadas de *LRRK2* (G2019S, R1441C o quinasa inactiva), presenta menor protección. Estos datos

sugieren que LRRK2 tiene un papel como protector de la disfunción mitocondrial y que algunos de sus mutantes podrían aumentar la vulnerabilidad neuronal a estreses asociados a la EP (Saha *et al.*, 2009).

No obstante, algunos estudios filogenéticos recientes afirman que los *LRRK* de *D. melanogaster* y *C. elegans* son ortólogos del *LRRK1*, y que por tanto estos organismos no son adecuados como modelos para determinar características de la proteína humana (Marín *et al.*, 2008). Dada la posibilidad de que los genes *LRRK* no sean verdaderos ortólogos de *LRRK2*, de la escasa equivalencia entre el sistema nervioso de los invertebrados con el humano y su corto periodo de generación que dificulta el análisis del efecto a edades tardías, resulta fundamental para una comprensión completa de la implicación de LRRK2 en la patofisiología de la EP el establecimiento y la caracterización detallada de modelos mamíferos. Los estudios realizados con modelos de ratón *knock out* y transgénicos para LRRK2, sugieren que esta proteína no es esencial para el desarrollo y la diferenciación neural y que no está implicada en rutas de muerte celular, al menos hasta la vejez de los animales. Sin embargo, estos modelos proporcionan valiosas herramientas para el mejor entendimiento de las funciones neuronales específicas de LRRK2 (Melrose, 2008; Yue, 2009).

#### **1.2.4. Función e interacciones con otras proteínas**

Aunque actualmente la función de dardarina continúa sin estar del todo clara, si se conocen algunos aspectos parciales de la misma. Se han realizado diversos estudios para identificar las proteínas que interaccionan con LRRK2, que además de ser útiles para encontrar nuevas dianas del estudio molecular de la EP, podrían arrojar luz sobre la función de esta proteína. La mayoría de las proteínas identificadas mediante estos métodos están relacionadas con el citoesqueleto o con el tráfico celular (revisado en Greggio y Cookson, 2009; Santpere y Ferrer, 2009).

LRRK2 interacciona *in vitro* con miosina, ezrina y radixina, proteínas encargadas del anclaje de la actina a la membrana plasmática. Además, el dominio ROC interacciona con los heterodímeros de  $\alpha/\beta$ -tubulina en los microtúbulos, colocalizando con ellos en las neuronas primarias del hipocampo (Greggio y Cookson, 2009). También se ha visto la interacción con proteínas involucradas en el proceso de

endocitosis, entre las que destacan la vimentina, la clatrina y la proteína RAB5, colocalizando además con esta última en las regiones presinápticas. Estos estudios junto con el hecho de que dardarina se asocia a diversas estructuras de membrana en el interior celular, indican que la función fisiológica de LRRK2 puede estar relacionada con la regulación del tráfico vesicular o incluso de las funciones sinápticas, que podría modular a través de la regulación de la endocitosis de las vesículas sinápticas (Shin *et al.*, 2008). Se ha observado que tanto la sobreexpresión como la disrupción de LRRK2 disminuyen los ratios de endocitosis de vesículas sinápticas en un cultivo celular de neuronas de hipocampo (Shin *et al.*, 2008). Además, otros estudios muestran que el efecto tóxico de la sobreexpresión de LRRK2 en neuronas podría deberse a los defectos endocíticos mostrados. Sin embargo, no se observan diferencias significativas en la cinética de la endocitosis al sobreexpresar LRRK2 silvestre y los mutantes típicos de la EP, por lo que posiblemente la función de regular el proceso de endocitosis no esté directamente relacionada con la patogénesis de la enfermedad (Shin *et al.*, 2008), aunque curiosamente los defectos en el tráfico proteico están comúnmente relacionados con varias enfermedades neurodegenerativas (Biskup *et al.*, 2006). Además, las interacciones con elementos del citoesqueleto se asocian a la regulación de la morfología y la motilidad celular, fundamentales en el desarrollo del sistema nervioso, y se ha descrito el papel de LRRK2 como regulador en la morfogénesis de las neuritas (Häbig *et al.*, 2008).

Sin embargo LRRK2 también se ha asociado a muchos otros procesos celulares, como por ejemplo, la regulación del inicio de la traducción proteica en situaciones de estrés a través de su interacción con el factor 4E-BP y el mantenimiento de la función mitocondrial (Greggio y Cookson, 2009) que juega un papel fundamental en la patogénesis de muchas enfermedades neurodegenerativas. Los procesos de fisión y fusión de la mitocondria están regulados entre otras por proteínas con dominios GTPasa y WD40. Dado que LRRK2 combina ambos dominios y que se presenta asociada a la membrana mitocondrial, esta proteína podría participar en el andamiaje en los eventos de fisión y fusión en este orgánulo (Biskup *et al.*, 2006). Por otra parte, la monitorización de los cambios en los patrones de expresión génica en células SH-SY5Y (línea dopaminérgica de neuroblastoma) provocados al disminuir los niveles de LRRK2,

también ha permitido relacionar LRRK2 con el crecimiento y la diferenciación celular, así como con la progresión del ciclo celular. Esta disminución de los niveles de LRRK2 también afecta a las cascadas de señalización de p53, una proteína encargada de la parada del ciclo celular y de la inducción de apoptosis (Häbig *et al.*, 2008). Además, se ha descrito la implicación de LRRK2 en la activación de la apoptosis debido a su interacción con FADD (*Fas-associated protein with death domain*), receptor implicado en un complejo de señales que resulta en la activación de la caspasa 8 y por lo tanto de la caspasa 3, tanto por la vía apoptótica extrínseca como intrínseca (Webber y West, 2009).

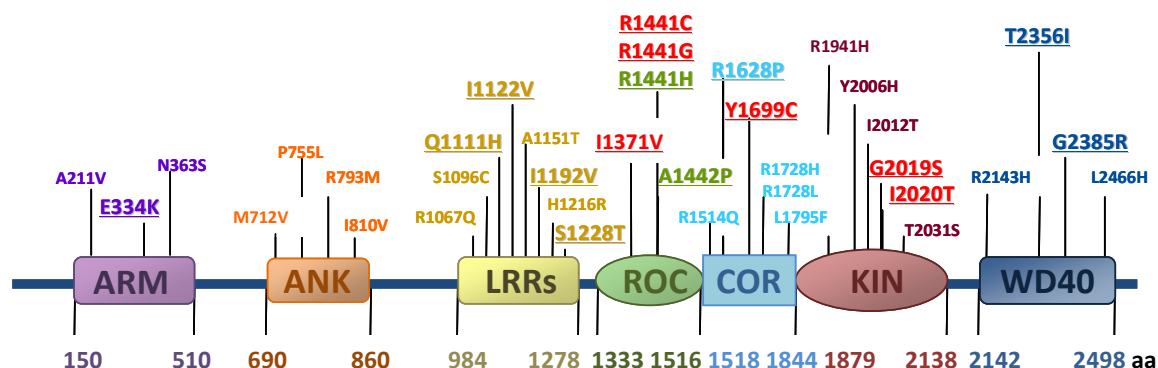
LRRK2 también interacciona con algunas chaperonas moleculares como Hsp90 y Cdc37, que parecen desempeñar un papel crucial en su estabilidad (Dächsel *et al.*, 2007). Además se han realizado estudios de interacción física con las principales proteínas involucradas en enfermedades neurodegenerativas, sin encontrarse para  $\alpha$ -sinucleína, tau o DJ-1, y sí para parkina, una E3 ubiquitina ligasa asociada a la EP; aunque no hay evidencias que esta proteína promueva la ubiquitinación o degradación de LRRK2 (Smith *et al.*, 2005). Cabe destacar que varias de las proteínas con las que interacciona dardarina son sustrato de fosforilación de su dominio quinasa, por ejemplo la miosina, ezrina, radixina y el factor 4E-BP (Jaleel *et al.*, 2007; Greggio *et al.*, 2009). En conclusión, los estudios realizados permiten postular que LRRK2 es una proteína compleja multidominio con relevancia en los mecanismos de señalización celular.

#### **1.2.5. Mutaciones de LRRK2 asociadas a la EP**

Se conocen al menos 20 mutaciones en LRRK2 ligadas a la EP con herencia autosómica dominante. Las mutaciones G2019S (dominio quinasa), R1441G/C (dominio GTPasa) e Y1699C (dominio COR) entre otras, además de segregar con la enfermedad en las formas familiares, han sido descritas como causantes de la EP de tipo esporádico. La EP ligada a mutaciones en LRRK2 provoca una progresión clínica y un fenotipo neuroquímico indistinguible de la patología típica de inicio tardío (West *et al.*, 2007). A pesar de que muchas de estas mutaciones causan un parkinsonismo

clínicamente típico, la neuropatología de la enfermedad varía notablemente entre pacientes, por ejemplo, en la presencia o no de cuerpos de Lewy (Tong *et al.*, 2009).

Se han encontrado mutaciones en LRRK2 en un gran número de pacientes con EP, de hecho en estos pacientes las mutaciones de *LRRK2* son tan frecuentes como las mutaciones de *TP53* en tumores sólidos (West *et al.*, 2007) y son las más frecuentes en poblaciones caucásicas (Luzón-Toro *et al.*, 2007). La mutación G2019S en particular, ha sido detectada en el 5-6 % de la EP familiar autosómica dominante, y en el 1-2 % de los casos esporádicos. Esta prevalencia ha sido superada en poblaciones específicas como en los judíos *Ashkenazi* y en los árabes del norte de África. Las mutaciones en LRRK2 se encuentran mayoritariamente en el extremo C-terminal, sugiriendo que este extremo es más importante a nivel funcional. No obstante, también se han encontrado sustituciones de aminoácidos en la región N-terminal a lo largo de los dominios ARM, ANK y LRR, y se postula que las mutaciones en estos dominios podrían interferir con la unión de dardarina a proteínas o moléculas, al igual que ocurriría con las mutaciones en el dominio WD40 (C-terminal) (Shin *et al.*, 2008; Mata *et al.*, 2006).



**Figura 1.5. Representación esquemática de la estructura de dominios de LRRK2 indicando las sustituciones de aminoácidos descritas para cada dominio.** Se indican subrayadas las mutaciones patogénicas asociadas con la EP y en rojo las que de éstas segregan con la enfermedad en familias, siendo el resto polimorfismos cuya implicación en la patogénesis de la EP no ha sido demostrada.

Sin embargo, los cambios de aminoácido más estudiados son los relacionados con los dominios catalíticos de la proteína. Como ya se ha comentado, la mutación más común es la G2019S, que provoca una hiperactividad quinasa, y por tanto un aumento de la autofosforilación y de la fosforilación de sustratos genéricos (Luzón-Toro *et al.*, 2007). La actividad catalítica está regulada por la autofosforilación en dos sitios del



segmento de activación, pero el cambio G2019S genera un sitio adicional de autofosforilación y dirige un cambio conformacional que mimetiza el estado activo de la quinasa (Greggio y Cookson, 2009). Este mecanismo puede explicar la ausencia de diferencias entre el fenotipo heterocigoto y homocigoto para la mutación G2019S, ya que las moléculas mutantes podrían fosforilar a las silvestres (Luzón-Toro *et al.*, 2007).

Otras mutaciones comúnmente encontradas en pacientes con EP son R1441G/C, que afectan al dominio GTPasa y parecen disminuir la actividad GTPasa pero aumentando la actividad quinasa (Santpere y Ferrer, 2009). Sin embargo la mutación R1441H parece inhibir la capacidad de LRRK2 de autofosforilarse, así como de fosforilar a MBP (*myelin basic protein*) y miosina. Las distintas sustituciones patogénicas de este aminoácido, que tridimensionalmente se encuentra accesible al solvente y distante al sitio de hidrólisis del GTP, sugieren que el mecanismo patogénico puede ser la pérdida de interacción con otras proteínas. Estos estudios apoyan la hipótesis que el dominio GTPasa de dardarina regula la actividad de su dominio quinasa (Mata *et al.*, 2006). También se ha descrito la mutación G2385R del dominio WD40 como atenuadora de la actividad quinasa (Jaleel *et al.*, 2007). No obstante, la influencia de las distintas sustituciones de aminoácidos en la actividad quinasa de la proteína es un tema controvertido, y los mecanismos por los que mutaciones de LRRK2 causan EP parecen ser complejos y no explicables únicamente por un incremento en la actividad quinasa.

#### 1.2.5.1. Efectos de las mutaciones de LRRK2 a nivel celular

El fenotipo clínico comúnmente asociado a las mutaciones de dardarina es una EP de inicio tardío, por lo que normalmente esta patología resulta indistinguible de la EP idiopática. Además, la neuropatología asociada a mutaciones de LRRK2 es variable, por lo que se sugiere que esta proteína está implicada en distintas rutas celulares. Se sabe que su afectación causa la pérdida de las neuronas dopaminérgicas y la acumulación anormal de proteínas ubiquitinadas fundamentalmente  $\alpha$ -sinucleína pero ocasionalmente tau hiperfosforilada (Santpere y Ferrer, 2009).

Se ha descrito que mutaciones patogénicas de LRRK2 como R1441C, Y1699C o G2019S, provocan degeneración neuronal, así como agregados proteicos o inclusiones

intracitoplasmáticas, en células SH-SY5Y y en cultivos de neuronas corticales (West *et al.*, 2005) y también incrementan los ratios de apoptosis al favorecer la interacción de LRRK2 con FADD (Webber y West, 2009). En general, la sobreexpresión de mutantes de LRRK2 provoca una importante pérdida de viabilidad, mientras que la inactivación de la quinasa de LRRK2 reduce esta toxicidad (West *et al.*, 2007). La correcta actividad quinasa es necesaria para que se den estos efectos, pero el fenotipo de formación de cuerpos de inclusión parece deberse al desplegado parcial de la proteína causado por la mutación Y1699C del dominio COR, aunque también depende en cierta medida de la actividad quinasa. Así pues, al reemplazar el dominio quinasa por otro inactivo, se bloquea la formación de cuerpos de inclusión y se retrasa considerablemente la muerte celular (Greggio *et al.*, 2006).

### **1.3. JUSTIFICACIÓN Y OBJETIVOS**

La EP es una de las patologías neurodegenerativas más importantes, en cuanto prevalencia y pérdida de calidad de vida de las personas que la padecen. Como ya se ha comentado, es una patología compleja que presenta un factor genético importante, con varios genes ligados a la patogénesis de la enfermedad. Uno de estos genes, *LRRK2*, codifica la proteína dardarina, que presenta algunas mutaciones implicadas en casos de EP autosómica dominante y en casos esporádicos. Dardarina es una proteína grande y compleja, que engloba siete dominios que incluyen dos actividades enzimáticas, dejando entrever su importancia a nivel celular. Sin embargo, ni su estructura molecular (excepto en el caso del dominio ROC) ni su función han sido esclarecidas, por lo que su estudio resulta de gran interés para el avance en el conocimiento de las bases moleculares de la enfermedad.

En este trabajo se pretende abordar el estudio de esta estructura, que además permitiría avanzar en el conocimiento de su función y regulación, y del efecto que sobre la misma puedan tener las distintas mutaciones descritas a lo largo de la proteína. Estos conocimientos también permitirían abordar la aplicabilidad de su uso como diana terapéutica, para la obtención y mejora de tratamientos contra la EP.

Otro objetivo del trabajo es conocer qué proteínas interaccionan con LRRK2, así como analizar su posible papel como proteínas relacionadas con los mecanismos

implicados en la aparición de la enfermedad. Estos resultados podrían servir, además de para avanzar en el conocimiento de la enfermedad, para su aplicación en el diagnóstico molecular de la EP.



## 2. MATERIALES Y MÉTODOS

### 2.1. Material biológico

#### 2.1.1. Cepas bacterianas

Para las clonaciones se utilizó la cepa de *Escherichia coli* DH5 $\alpha$  (F $\phi$ 80*lacZ* $\Delta$ M15  $\Delta$ (*lacZ*YA-argF) U169 *recA1 endA1 hsdR17* (*r<sub>k</sub><sup>-</sup>*, *m<sub>k</sub><sup>+</sup>*) *phoA supE44 thi-1 gyrA96 relA1 tonA*).

Para la expresión de proteínas en procariotas se utilizó la cepa de *E. coli* BL 21-CodonPlus (DE3)-RIL (*E. coli* B F<sup>-</sup> *ompT hsdS* (*r<sub>B</sub><sup>-</sup>* *m<sub>B</sub><sup>-</sup>*) *dcm* + Tet<sup>r</sup> *E. coli* gal  $\lambda$  (DE3) *endA Hte* [*argU ileY leuW Cam<sup>r</sup>*]), que contiene un plásmido con resistencia a cloranfenicol que codifica los ARNt necesarios para la correcta traducción de proteínas heterólogas.

#### 2.1.2. Plásmidos utilizados

Tabla 2.1. Nomenclatura y descripción de los plásmidos utilizados durante el trabajo

PLÁSMIDO	DESCRIPCIÓN
pPROEX-HTa	Vector de expresión en procariotas. Amp <sup>R</sup> . Expresión heteróloga de proteínas fusionadas a 6His.
pPROEX-LRR-LRRK2	Vector pPROEX-HTa con el cDNA codificante del dominio LRR (aas 979-1310) de LRRK2.
pPROEX-COR-LRRK2	Vector pPROEX-HTa con el cDNA codificante del dominio COR (aas 1513-1745) de LRRK2.
pPROEX-ROCO-LRRK2	Vector pPROEX-HTa con el cDNA codificante los dominios ROC-COR (aas 1336-1745) de LRRK2.
pPROEX-KIN-LRRK2	Vector pPROEX-HTa con el cDNA codificante del dominio quinasa (KIN, aas 1876-2134) de LRRK2.
pPROEX-ROKOKIN-LRRK2	Vector pPROEX-HTa con el cDNA codificante de los dominios ROC-COR-KIN (aas 1336-2134) de LRRK2.

PLÁSMIDO	DESCRIPCIÓN
pPROEX-COR2-LRRK2	Vector pPROEX-HTa con el cDNA codificante del dominio COR (aas 1518-1844) de LRRK2.
pPROEX-ROCO2-LRRK2	Vector pPROEX-HTa con el cDNA codificante del dominio ROC-COR (aas 1333-1844) de LRRK2.

### 2.1.3. Medios y condiciones de cultivo

En todos los casos se añaden los antibióticos necesarios al medio después de autoclavarlo (100 ng/L de ampicilina y/o 33 ng/L de cloranfenicol y/o 50 ng/L de kanamicina).

- LB: 10 g/L de triptona, 5 g/L de extracto de levadura y 10 g/L de NaCl
- LB-agar: LB conteniendo 15 g/L de agar
- SOB: 20 g/L de triptona, 5 g/L de extracto de levadura, 564mg/L de NaCl y 187 mg/L de KCl.
- SOC: SOB conteniendo 10 mM  $\text{MgCl}_2\text{-MgSO}_4$  y de 2 mM glucosa (añadidos después de autoclavar y esterilizados por filtración).
- Medio de autoinducción (para 1 L): 925ml de ZY + 1 mL de  $\text{MgSO}_4$  (1 M) + 20 mL de 50 x 5052 + 50 mL de 20 x NPS + 1 mL ampicilina + 1 mL cloranfenicol

-ZY (1 L): 10 g triptona + 5 g extracto de levadura + 925 mL agua (autoclavado)

-50x 5052 (1 L): 250 g glicerol + 25 g glucosa + 100 g  $\alpha$ -lactosa + 730 mL agua + (autoclavado y conservado a 4 °C)

-20xNPS (1 L): 66 g  $(\text{NH}_4)_2\text{SO}_4$  + 136 g  $\text{KH}_2\text{PO}_4$  +142 g  $\text{Na}_2\text{HPO}_4$  + 900 mL agua (filtrado)

La temperatura de incubación en medio líquido varía dependiendo de la condición de expresión utilizada, siendo en todos los casos la agitación de 180 rpm.

## **2.2. Métodos de manipulación y análisis de proteínas**

### **2.2.1. Geles de acrilamida y tinción de Coomassie**

Los geles de acrilamida para *SDS-PAGE*, el tampón de carga y la tinción de Coomassie utilizados para visualizar proteínas se realizaron según el procedimiento descrito en el *Current Protocols in Protein Science* (Gallagher, 1995). El marcador de pesos moleculares (MPM) empleado es PageGel Two-Color SDS<sup>TM</sup> (Ref. EM 05160, *PAGE-gel*).

### **2.2.2. Cuantificación de proteínas**

Para la determinación de la concentración proteica se han utilizado dos métodos: la medida de Abs<sub>280</sub> y el método de Bradford (Olson y Markwell, 2007).

### **2.2.3. Expresión y obtención de proteínas**

A partir de un inóculo crecido a saturación de bacterias de *E. coli* BL 21-CodonPlus (DE3)-RIL se crecen 15 mL de LB o de medio autoinducible. Para el caso del LB se induce la expresión con IPTG (concentraciones desde 1 mM a 0,1 mM dependiendo del tiempo y la temperatura de inducción empleados), en el caso del medio autoinducible la inducción se da cuando se agota la glucosa del medio. Se recogen las células por centrifugación a 12.000 g durante 5 min a 4 °C y se descarta el sobrenadante. El sedimento se resuspende en el tampón de extracción de proteínas (100 mM Tris-HCl, 0,5 M NaCl suplementado con inhibidor de proteasas PMSF o *Complete* (Roche)). Se sonicán las células con 10 pulsos de 10 s e intervalos de 45 s en hielo, y se centrifuga a 12.000 g durante 30 min a 4 °C. El sedimento obtenido se resuspende en un volumen equivalente al del sobrenadante, y ambas fracciones se visualizan en un gel de acrilamida para determinar la proporción de proteína soluble.

### **2.2.4. Obtención y purificación de cuerpos de inclusión**

Las células procedentes de 1,5 L de cultivo se resuspenden en 10 mL de PBS 1X (*Phosphate-Buffered Saline*; 137 mM NaCl, 2,7 mM KCl, 10 mM Na<sub>2</sub>HPO<sub>4</sub>·2H<sub>2</sub>O, 2 mM

KH<sub>2</sub>PO<sub>4</sub> a pH 7,4) suplementado con inhibidor de proteasas y se sonica con 15 pulsos de 45 s intercalados con intervalos de 45 s en hielo. Posteriormente se centrifuga a 12.000 g durante 30 min a 4 °C y se descarta el sobrenadante, resuspendiendo el sedimento en 20 mL de la solución de lavado (50 mM Tris-HCl pH 8, 100 mM NaCl, 0,5 % Triton 100X y 0,1 % azida sódica). Se centrifuga a 15.000 g durante 20 min a 4 °C y se descarta el sobrenadante, resuspendiendo de nuevo el sedimento en 20 mL de solución de lavado mediante homogenización. Este procedimiento se repite 4 ó 5 veces hasta que se obtiene un sedimento de color blanco. Posteriormente se resuspende el sedimento en el mismo tampón sin Triton ni azida sódica y se vuelve a centrifugar a 15.000 g durante 20 min a 4 °C. El sedimento obtenido se introduce en 3-5 mL de una solución desnaturizante con urea (4-8 M urea, 50 mM Tris-HCl pH 8, 100 mM NaCl y 10 mM DTT) y se deja en agitación a 4 °C durante 24-48 h. Finalmente se centrifuga a 4.000 g durante 30 min a 4 °C y se recoge el sobrenadante, se mide la concentración proteica y se carga en un gel de acrilamida para determinar la pureza de la proteína de interés mediante *SDS-PAGE*.

### 2.2.5. Cromatografía de afinidad con resina de Ni

Para la purificación de proteínas con una cola de 6 His por cromatografía de afinidad se utilizó una resina de níquel-sefarosa 6 *fast-flow* de *GE Healthcare* (Ref. 17-5318-01).

Se lava la resina con 3 volúmenes de agua para eliminar el etanol de conservación y posteriormente se equilibra con otros 3 volúmenes del tampón donde está disuelta la proteína (tampón de desnaturalización). Posteriormente se añade la disolución de proteína desnaturizada y se lava con 3 volúmenes del tampón de equilibrado y con 2 volúmenes del mismo tampón conteniendo 20 mM de imidazol. La proteína se eluye con 3 volúmenes del tampón de equilibrado con 250 mM de imidazol y la resina se conserva en etanol al 20 %. Este proceso se realiza a 4 °C para evitar la degradación proteica.

También se utilizó un aparato de FPLC (*Fast Protein Liquid Chromatography*) *ÄKTA purifier* que permite la automatización del sistema y la elución utilizando un gradiente de imidazol. Para utilizar este sistema se requiere que la muestra sea



previamente filtrada y desgasificada. Las distintas fracciones recogidas de este procedimiento se cargaron en un gel de acrilamida para determinar la cantidad y pureza de la proteína en las diferentes fracciones.

### 2.2.6. Replegado proteico *in vitro*

Previo a la renaturalización, a la solución de la proteína desnaturalizada purificada por cromatografía se le añade 5-10 mM de DTT y se deja en agitación a 4 °C durante 1 h, para asegurar la rotura total de los puentes disulfuro.

Existe una extensa batería de tampones utilizados en la renaturalización *in vitro* de proteínas, ya que cada proteína requiere unas condiciones específicas. En la Tabla 2.2 se encuentra la composición de los tampones utilizados en este trabajo. En la primera parte (A) se muestra la composición de una batería de tampones de replegado de pH 6 mientras que los tampones de la segunda se muestran los de pH 8,5 (B).

Tabla 2.2.

(A)

Stock	T. 1	T.2	T.3	T. 4	T.5	T.6	T.7
MES (1000mM)	50	50	50	50	50	50	50
NaCl (5000mM)	9.6	9.6	9.6	240	240	240	240
KCl (500mM)	0.4	0.4	0.4	10	10	10	10
MgCl <sub>2</sub> (500mM)	2	2		2			2
CaCl <sub>2</sub> (500mM)	2	2		2			2
Gua-HCl (5000mM)	750		750	750	750		750
Triton 100X (100 %)	0.5		0.5			0.5	
PEG 3350 (5 %)		0.05	0.05			0.05	0.05
DTT (1000mM)	1		1		1		1
Arginina (1000mM)		500				500	
GSH/GSSG (1000/100mM)		1/0.1		1/0.1		1/0.1	
Sacarosa (2000mM)			200		200	200	
EDTA (1000mM)			1		1	1	

(B)

Stock	T.8	T.9	T.10	T.11	T.12	T.13	T.14	T.15
TRIS-HCl (1000mM)	50	50	50	50	50	50	50	50
NaCl (5000mM)	9.6	9.6	9.6	9.6	240	240	240	240
KCl (500mM)	0.4	0.4	0.4	0.4	10	10	10	10
MgCl <sub>2</sub> (500mM)	2		2				2	2
CaCl <sub>2</sub> (500mM)	2		2				2	2
Gua-HCl (5000mM)	750	750	750			750	750	
Triton 100X (100 %)				0.5	0.5		0.5	
PEG 3350 (5 %)		0.05			0.05		0.05	
DTT (1000mM)		1		1		1		1
Arginina (1000mM)	500	500	500			500	500	
GSH/GSSG (1000/100mM)	1/0.1		1/0.1		1/0.1		1/0.1	
Sacarosa (2000mM)	200		200				200	200
EDTA (1000mM)		1		1	1	1		1

Al tampón de renaturalización (suplementado con inhibidor de proteasas) en agitación lenta a 4 °C se le añade la solución proteica gota a gota, de forma que quede como mínimo diluida 20 veces, y se incuba en estas condiciones de 24 a 48 h. A las 24 h se puede añadir de nuevo más disolución de proteína para aumentar la eficiencia del proceso, siempre teniendo en cuenta que la concentración final de proteína no sea excesiva. Tras esto se centrifuga a 4.000 g durante 1 h a 4 °C para eliminar los posibles agregados proteicos, y se concentra la proteína soluble mediante centricones de Millipore (x24 AMICON ULTRA-1510K). Con este método además de concentrar la proteína también es posible sustituir el tampón por 100 mM Tris-HCl pH 8 y 0,5 M NaCl, que se va añadiendo progresivamente para sustituir al tampón de replegado.

### 2.2.7. Cromatografía de exclusión molecular

Para determinar si la proteína tras el replegado se encuentra en forma de monómero, dímero u oligómero se realizó una cromatografía de exclusión molecular

en un aparato ÄKTA purifier (GE Healthcare ) con una resina Superdex™ 200 10/300 gL de GE Healthcare.

Se lava la resina con 3 volúmenes de agua para eliminar el etanol de conservación residual y posteriormente se equilibra con otros 3 volúmenes del tampón donde está disuelta la proteína (100 mM Tris-HCl pH 8 y 0,5 M NaCl) filtrado y desgasificado. Posteriormente se calibra con diferentes sustancias de masa molecular (Mm) conocida, para generar una recta patrón con los volúmenes de retención obtenidos para cada una de ellas. Tras el calibrado se pasó la muestra proteica y se determinó su volumen de retención, calculándose la Mm de la proteína en base a la extrapolación en la recta patrón. Al finalizar se lava la columna en agua y se conserva en etanol al 20 %.

### 2.2.8. Prueba de estabilidad proteica

Este procedimiento simplifica la identificación de las condiciones óptimas en las que la proteína es estable, información fundamental para los ensayos de cristalización. Se basa en el uso de un colorante solvatocrómico (como el SYPRO Orange, BIO-RAD) como indicador de la desestructuración proteica, ya que la unión de este tipo de cromóforo a una proteína desplegada provoca un incremento de su fluorescencia intrínseca (Forneris *et al.*, 2009).

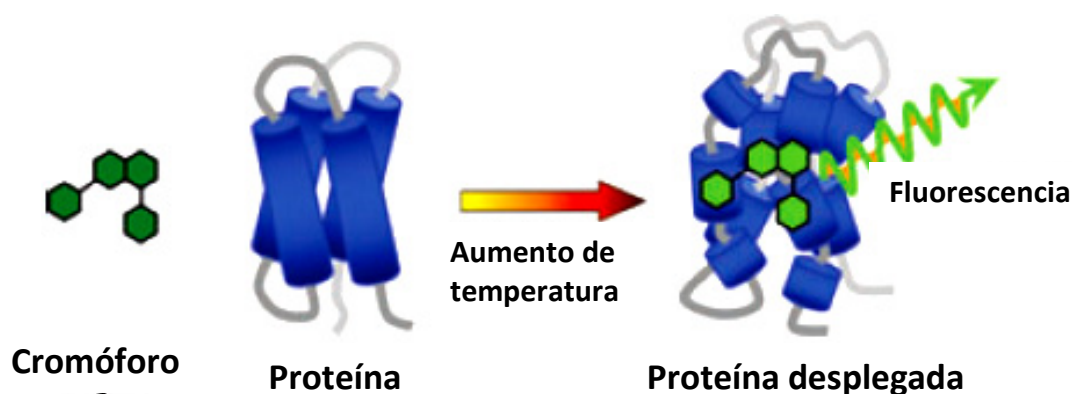


Figura 2.1. Representación del principio de la prueba de estabilidad proteica (modificada de Forneris *et al.*, 2009).

Previo al ensayo de estabilidad, se determina si se puede utilizar este ensayo (basado en fluorescencia) con la proteína de interés, así como qué concentración de proteína debe utilizarse. Para ello se realizan distintas diluciones, comprendidas entre

0 y 1 mg/mL de proteína, en un tampón estándar (100 mM HEPES, 150 mM NaCl, pH 7,5), se mezclan 40 µL de cada dilución con 1 µL del fluoróforo *Sypro Orange* a 200x y se alicuotan por duplicado 20 µL de la mezcla en una placa de 96 pocillos, sellada con un adhesivo ópticamente inactivo. La placa se introduce en un aparato 7500 *Fast Real-Time PCR* (*Applied Biosystems*), con un programa que abarca de 20 °C a 70 °C, aumentando 1 °C por ciclo. Se representa gráficamente la fluorescencia medida en cada ciclo y se obtiene la  $T_m$ , temperatura a la cual la mitad de la proteína se encuentra desnaturalizada, calculada según el modelo de Boltzmann.

Tras el ensayo para determinar la concentración a utilizar se realiza la prueba de estabilidad de la proteína de interés en 48 tampones distintos. A 36 µL de cada tampón se le añade 4 µL de una mezcla de la proteína con el fluoróforo, de forma que al final la proteína quede a la concentración óptima determinada anteriormente y el fluoróforo a 5x. Se alícuota por duplicado 20 µL en una placa de 96 pocillos y se realiza el mismo procedimiento que para el ensayo previo. La distribución y composición de los diferentes tampones utilizados se detalla en el Anexo 3.

#### **2.2.9. Obtención de extracto de cerebro de rata**

Después de lavar el cerebro de rata con 10 mL del tampón de extracción (100 mM Tris-HCl, 0,5M NaCl, 1 % Triton 100X, 10 mM EDTA e inhibidor de proteasa) se sumerge en 10 mL del mismo tampón y se homogeniza utilizando un homogenizador de *IKA LABORTECHNIK* (*Ultra Turbax T25*). El homogenizado se centrifuga a 12.000 g durante 45 min a 4 °C y el sobrenadante se filtra con un filtro de 0,2 µm de poro. El extracto resultante se cuantifica mediante el método de Bradford y se visualiza en un gel de acrilamida para comprobar la integridad de las proteínas.

#### **2.2.10. Ensayos de pull-down**

Para los ensayos de pull-down se utilizó la misma resina que para la cromatografía de afinidad. Tras lavar la resina con 3 volúmenes de agua para eliminar el etanol residual, se equilibra con 3 volúmenes del tampón de disolución de proteínas. Posteriormente se introduce la proteína de interés (entre 6-8 mg) y se lava con 3 volúmenes del tampón de equilibrado y con 2 volúmenes del mismo tampón con 20 mM de imidazol. Tras volver a equilibrar la resina con el tampón de equilibrado se

pasan 75 mg de extracto de proteína de cerebro de rata, se lava con 5 volúmenes de tampón de equilibrado y con 2 volúmenes del mismo tampón con 20 mM de imidazol. Finalmente, se eluyen las proteínas retenidas en la resina con un lavado de 3 volúmenes del tampón de equilibrado con 250 mM de imidazol y la resina se conserva en etanol al 20 %. Las fracciones de interés recogidas se visualizan en un gel de acrilamida. La tinción del gel resultante se realiza con Coomassie sin acético, para poder secuenciar las proteínas a partir de las bandas en caso de ser de interés. Este procedimiento, al igual que la cromatografía de afinidad, se realiza a 4 °C para evitar la degradación proteica.

#### **2.2.11. Concentración de proteínas por precipitación con acetona**

Para concentrar las disoluciones obtenidas del ensayo de *pull-down* se precipitan las proteínas añadiendo 10 volúmenes de acetona a la muestra e incubando la mezcla durante 1 h a -80 °C. Posteriormente se centrifuga a 4.000 g durante 20 min a 4 °C y se descarta el sobrenadante, resuspendiendo el sedimento en el volumen deseado del tampón correspondiente.

#### **2.2.12. Análisis *in silico* de la estructura de los dominios de LRRK2**

Se analiza la estructura de cada dominio según las predicciones obtenidas para la estructura secundaria y terciaria. La predicción tridimensional de los dominios se realiza por homología de secuencia (en el caso de que exista una proteína de estructura resuelta con una homología de secuencia del 30 % como mínimo) o por *threading* (en el caso de no existir moldes adecuados). También se estudian los posibles cambios de estructura debidos a las mutaciones patogénicas asociadas a la EP y algunos de los polimorfismos descritos, tanto realizando una nueva predicción de la estructura de la proteína mutante como introduciendo la sustitución de aminoácido con el programa de visualización sobre la predicción de la proteína silvestre. En el apartado 2.4 se hace referencia a las herramientas bioinformáticas utilizadas para el análisis *in silico*.

## 2.3. Métodos de manipulación y análisis de ácidos nucleicos

### 2.3.1. Electroforesis de DNA en geles de agarosa

Los geles de agarosa para la visualización de ácidos nucleicos y el tampón de carga se realizaron según el procedimiento descrito por *Current Protocols in Molecular Biology* (Voytas, 2000). El marcador de pesos moleculares (MPM) empleado es *1kb Plus DNA Ladder* (Ref. 10787026, *Invitrogen*).

### 2.3.2. Amplificación y preparación del inserto

En primer lugar, se diseñan cebadores flanqueantes de los dominios de interés, COR2 (aas 1518-1844: F\_ACACCATGGGTTTCAAGATCCGAG y R\_TCTAGACTAGTTCAAATTCCAACACA) y ROCO2 (aas 1333-1844: F\_ACAGTCGACGTATGAACTTATGATTG y el mismo reverso que para COR), incluyendo algunos nucleótidos para generar los sitios de restricción adecuados. Se amplifica el inserto por PCR utilizando 0,1 mM de dNTPs, 0,4 µM de cebadores, 2,5 U de *Pfu* de *Stratagene* con su respectivo tampón, 50 ng de DNA plasmídico con el cDNA de *LRRK2* completo y el resto hasta 50 µL de agua miliQ autoclavada. Las condiciones utilizadas para la amplificación de los insertos son 2 min a 95 °C, seguido de 35 ciclos de 1 min a 95 °C, 30 s a 50 °C y 1 min 30 s (COR2) o 2 min (ROCO2) a 72 °C, y finaliza con 10 min a 72 °C de extensión adicional. Se realizan dos amplificaciones por inserto para aumentar la cantidad del mismo en la reacción de ligación. El producto de la PCR se carga en un gel del 0,8 % de agarosa para comprobar la especificidad de la reacción y el tamaño del producto (1225 pb para COR2 y 1665 pb para ROCO2). Tras esta comprobación los productos de PCR para el mismo inserto se juntan y se concentran por precipitación, añadiendo al producto de PCR 2 µL de glucógeno a 20 mg/mL, 20 µL de acetato amónico 10 M y 250 µL de isopropanol. La mezcla resultante se deja durante 1 h a -80 °C u o/n (*over night*) a -20 °C, se centrifuga durante 10 min a 12.000 g, se lava el sedimento con etanol al 70 % y se vuelve a centrifugar durante 10 min a 12.000 g. Finalmente, el sedimento se resuspende en 16 µL de agua miliQ autoclavada.

Se realiza la digestión de los insertos en un volumen de 20 µL con 10 unidades de cada enzima (*Nco* I y *Xba* I para COR2, *Sal* I y *Xba* I para ROCO2, todos ellos de *Fermentas*) e incubando la mezcla a 37 °C o/n. Se digiere 100 ng del vector pPROEX-

HTa con los mismos enzimas y tras la inactivación enzimática (10 min a 65 °C) se purifica el producto de la restricción mediante precipitación y se resuspende en 10 µL. Para realizar un cálculo aproximado de las concentraciones de inserto y vector y así ajustar las cantidades a añadir de cada uno para la ligación, se cargan en un gel del 0,8 % de agarosa 1 µL de cada uno junto al marcador de pesos moleculares *Real Ladder 1* (Ref. RBMM1, Durviz) que permite cuantificar aproximadamente la cantidad de DNA de las bandas.

### 2.3.3. Producción de bacterias *E. coli* DH5α electrocompetentes

Se pica una colonia de *E. coli* DH5α y se crece durante el día en 1 mL de LB, después se plaquean 100 µL del cultivo en sendas placas de LB-ampicilina LB-kanamicina. El resto del cultivo se crece o/n en 40 mL de SOC para generar el inóculo. El día siguiente, tras comprobar que no han crecido colonias en las placas, se inocula 1 L de SOC con 10 mL del inóculo y se crece a 37 °C hasta una DO<sub>595</sub> de 0,5-0,6. Después se enfría en hielo y se centrifuga a 4.000 g durante 10 min a 4 °C, se lava el sedimento de células con 1 L de agua miliQ autoclavada y se vuelve a centrifugar. Este proceso se repite dos veces lavando con 0,5 L de agua miliQ. El sedimento de células resultante se resuspende en 20 mL de glicerol 10 % y se centrifuga a 12.000 g durante 20 min a 4 °C. El sedimento se resuspende en 2-3 mL de glicerol 10 % y se separa en alícuotas 40 µL en tubos que se congelan inmediatamente en N<sub>2</sub> líquido y se conservan a -80 °C. El material utilizado (puntas, tubos, etc.) en este protocolo se enfría previamente a -20 °C. Finalmente, se comprueba mediante siembra en placas de ampicilina y kanamicina que no existe crecimiento (bacterias contaminantes resistentes a antibióticos).

### 2.3.4. Ligación y transformación

La desfosforilación del vector y la ligación se realiza mediante el *Rapid DNA Dephos and Ligation Kit* (Roche, Ref. 04898117001) y el producto de la ligación se purifica mediante *High Pure PCR Product Purification Kit* (Ref. 11732676001, Roche), en ambos casos siguiendo las instrucciones del fabricante. Posteriormente se transforma 40 µL de bacterias electrocompetentes *E. coli* DH5α con 10 µL del producto de ligación mediante electroporación, aplicando un pulso eléctrico de 1.700 mV. Las bacterias se resuspenden rápidamente en 200 µL de SOC y se incuban 45 min en

agitación a 37 °C. Posteriormente se siembran en una placa de LB agar con ampicilina y se incuba a 37 °C o/n.

### **2.3.5. Comprobación de la ligación, extracción del DNA plasmídico y obtención de los clones**

Se verifica si las colonias resultantes han incorporado el plásmido con el inserto mediante la amplificación por PCR de colonia utilizando cebadores presentes en el plásmido (F\_GAGCGGATAACAATTTACACAGG y R\_GGCTGAAAATCTTCTCTCATCCG), de forma que se obtiene un fragmento de unas 200 pb si el plásmido está vacío, y de 200 pb más el tamaño del inserto si ha incorporado éste. En este caso la reacción de PCR contiene 75 µM de dNTPs, 0,3 µM de cebadores, 0,1 U de Taq polimerasa de *Qiagen* con su respectivo tampón, una colonia bacteriana y el resto hasta 10 µL de agua miliQ autoclavada. Las condiciones para la amplificación son 10 min a 95 °C; 35 ciclos de 30 s a 95 °C, 30 s a 60 °C y 1 min 30 s (COR2) o 2 min (ROCO2) 72°C y 10 min finales a 72°C. Los productos de PCR se cargan en un gel de agarosa y de aquellas colonias que sean positivas se extrae el DNA plasmídico utilizando el método descrito en *Current Protocols in Molecular Biology* (Engebrecht *et al.*, 1991), y se secuencian para comprobar que no existe ningún cambio en la secuencia de DNA. Posteriormente se transforma con el plásmido obtenido la cepa *E. coli* BL 21-CodonPlus (DE3)-RIL.

### **2.4. Herramientas bioinformáticas**

Para la predicción de la masa molecular, el punto isoeléctrico y el coeficiente de extinción de las proteínas se utilizó el *ProtParam tool* del servidor ExPASy *proteomics tools* (Gasteiger *et al.*, 2005; <http://www.expasy.ch/tools/protparam.html>). Para la predicción de la estructura secundaria se utilizó JPRED3 (Cole *et al.*, 2008; <http://www.compbio.dundee.ac.uk/www-jpred/>), y para la terciaria se utilizaron Swiss-model (modelado por homología; Schwede *et al.*, 2003; <http://swissmodel.expasy.org//SWISS-MODEL.html>) y LOMETS (modelado por *threading*; Wu y Zhang, 2007; <http://zhang.bioinformatics.ku.edu/LOMETS/>). Para la visualización de estructuras tridimensionales de proteínas se utilizó PyMOL (DeLano Scientific Inc, 2002). Para el diseño y evaluación de los cebadores se utilizó el programa *Gene Runner* (Hastings software Inc, 1994).



### 3. RESULTADOS Y DISCUSIÓN

#### 3.1. Trabajo con dominios proteicos de LRRK2

En este apartado se ha trabajado con las definiciones de dominios descritas en el apartado 2.1.2 de Materiales y Métodos, que habían sido previamente clonadas en el laboratorio en el que se ha realizado el proyecto.

##### 3.1.1. Determinación de las características físico-químicas

En primer lugar, mediante la herramienta *ProtParam* se predijeron los principales parámetros físico-químicos de los dominios como la masa molecular (Mm), el punto isoeléctrico (pI) y el coeficiente de extinción molar ( $\epsilon$ ). El conocimiento de estos parámetros es de utilidad para la visualización de los dominios en geles de acrilamida, y para la extracción y la cuantificación de proteínas, respectivamente.

Tabla 3.1. Parámetros físico-químicos de las construcciones empleadas.

Construcción	Mm (Da)	pI	$\epsilon$ (M <sup>-1</sup> cm <sup>-1</sup> )	
			Todas las Cys forman puentes S-S	Ninguna Cys forma puentes S-S
LRR	37918,8	5,91	25940	25440
COR	27687,5	8,6	37150	36900
ROCO	47813,8	9,28	59610	59360
ROCKIN	91472,3	6,77	96105	95230
KIN	29323,9	6,5	22140	21890

Se especifica para cada dominio su masa molecular relativa, punto isoeléctrico y coeficiente de extinción a Abs<sub>280</sub> medido en agua. Este último se especifica en la situación de que todas las cisteínas o ninguna se encuentren formando puentes disulfuro, empleándose para los cálculos posteriores la media de ambos.

##### 3.1.2. Expresión y pruebas de solubilidad

Se probaron diferentes condiciones de expresión variando la temperatura y el tiempo de crecimiento, utilizando tanto medio LB e inducción de la expresión mediante IPTG como medio autoinducible. Estas condiciones se esquematizan en la Tabla 3.2.

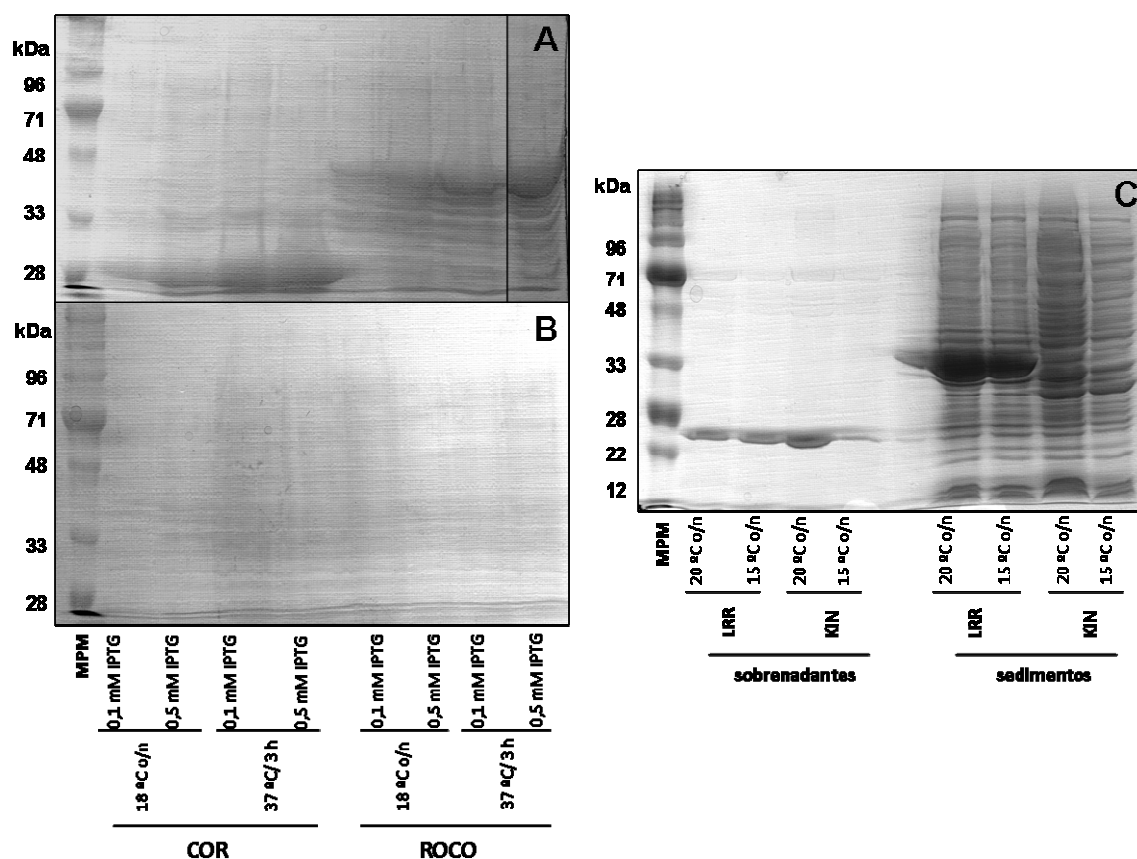
Tabla 3.2. Condiciones de crecimiento y expresión ensayadas en las pruebas de solubilidad.

	Crecimiento	Inducción-Expresión		Construcciones
Medio LB	3 h a 37 °C hasta DO <sub>595</sub> entre 0,3 y 0,5	IPTG 1 mM	37 °C / 3 h	COR, ROCO, ROKOKIN, LRRs y KIN
			30 °C / 5 h	COR, ROCO y ROKOKIN
			25 °C o/n	COR, ROCO, ROKOKIN, LRRs y KIN
			20 °C o/n	COR, ROCO y ROKOKIN
			15 °C o/n	
		IPTG 0,5 mM	37 °C / 3h	
			18 °C o/n	
		IPTG 0,1 mM	37 °C / 3h	
			18 °C o/n	
		20 °C hasta DO <sub>595</sub> de 0,6	IPTG 1 mM	20 °C o/n
15 °C hasta DO <sub>595</sub> de 0,6	15 °C o/n	COR, LRRs y KIN		
Medio Autoinducible	a 37 °C hasta una DO <sub>595</sub> de 1	25 °C o/n		COR, ROCO y ROKOKIN
		20 °C o/n		
		15 °C o/n		

Se representa el medio utilizado y las condiciones de crecimiento, inducción y expresión; indicando en qué dominios se ha ensayado cada una de las condiciones.

Como se indica en la tabla 3.2 con el objetivo de conseguir proteína soluble se utilizaron diversas condiciones de expresión, consiguiéndose una expresión considerable de la proteína en todos los casos. Entre las condiciones ensayadas, se utilizaron algunas de inducción suave, reduciendo la temperatura y la concentración

de IPTG y aumentado el tiempo, o utilizando el medio de autoinducción, sin conseguirse en ningún caso proteína soluble (Figura 3.1). La ausencia de solubilidad podría deberse a una definición errónea de los dominios, o al uso de un modelo procariota para la expresión, ya que entre procariotas y eucariotas existen diferencias importantes en los mecanismos de plegamiento proteico y las modificaciones postraduccionales.



**Figura 3.1. Pruebas de expresión y solubilidad de COR, ROCO, LRR y KIN.** A) Muestra los sedimentos de las pruebas de las construcciones de COR y ROCO. B) Muestra los respectivos sobrenadantes de las pruebas representadas en el gel A. C) Muestra las pruebas de expresión y solubilidad con 1 mM de IPTG de los dominios LRR y KIN. El tampón de extracción utilizado para todos los casos es de pH 7,5. Cabe destacar que se observa mayor expresión proteica a 37 °C.

El pH del tampón de extracción de proteínas utilizado en la extracción varía según el pI de la construcción estudiada, probándose en cualquier caso varios pH para cada construcción con el objetivo de aumentar las posibilidades de obtener proteína soluble. Se probaron los pH 6,5, 7,5 y 8,5 para las construcciones LRR y KIN, y además

el pH 7,5 se ensayó para COR y ROCO y el pH 8,5 para COR y ROCOKIN. A pesar de ello, no se obtuvo proteína soluble para ninguna de las condiciones ensayadas.

### 3.1.3. Replegado *in vitro*

Dado que no se consiguió obtener proteína soluble pero que los niveles de expresión eran elevados, se procedió al replegado *in vitro* de algunos de los dominios. Se planteó como más interesante la estructura del dominio COR, dado que, como se ha comentado en la introducción, no se conoce demasiado acerca de su función. Por estos motivos, se realizó el replegado *in vitro* de los dominios COR y ROCO. Se descartó el replegado de la combinación de dominios ROCOKIN, debido a su elevado tamaño.

Con el fin de obtener proteína en cantidad suficiente, se expresaron ambas construcciones en 1,5 L de cultivo, en las condiciones en las que se había observado una mayor expresión (crecimiento 37 °C hasta una  $DO_{595}$  de 0,6, e inducción con 1 mM de IPTG 3 h/37 °C). Tras la obtención de las células producidas en estas condiciones se purificaron los cuerpos de inclusión (véase Materiales y Métodos apartado 2.2.4). Los cuerpos de inclusión se disolvieron en tampones de desnaturalización con distinta concentración de urea (4 M, 6 M y 8 M), con el objetivo de determinar la mínima concentración de urea necesaria para desnaturalizar la máxima cantidad posible de proteína. De esta forma se determinó que la concentración óptima de urea del tampón de desnaturalización era de 6 M (Figura 3.2), con la cual se consigue una cantidad considerable de proteína soluble.

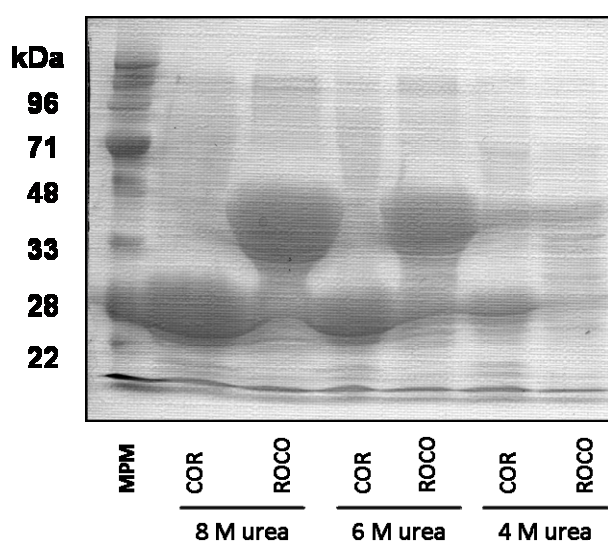
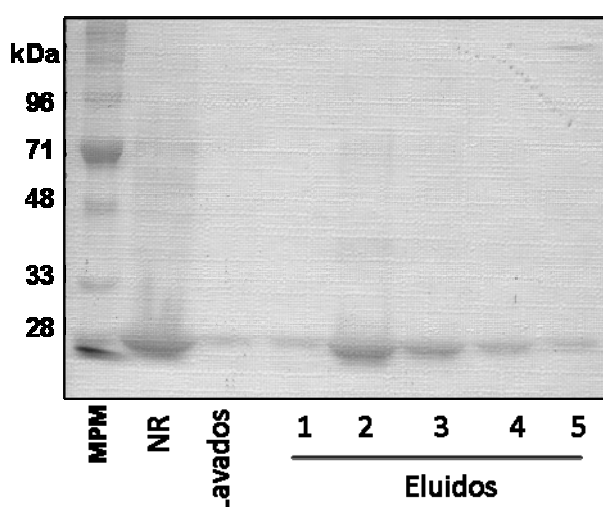


Figura 3.2. Prueba de condiciones de desnaturalización en tampón urea de los cuerpos de inclusión obtenidos para las construcciones COR y ROCO.

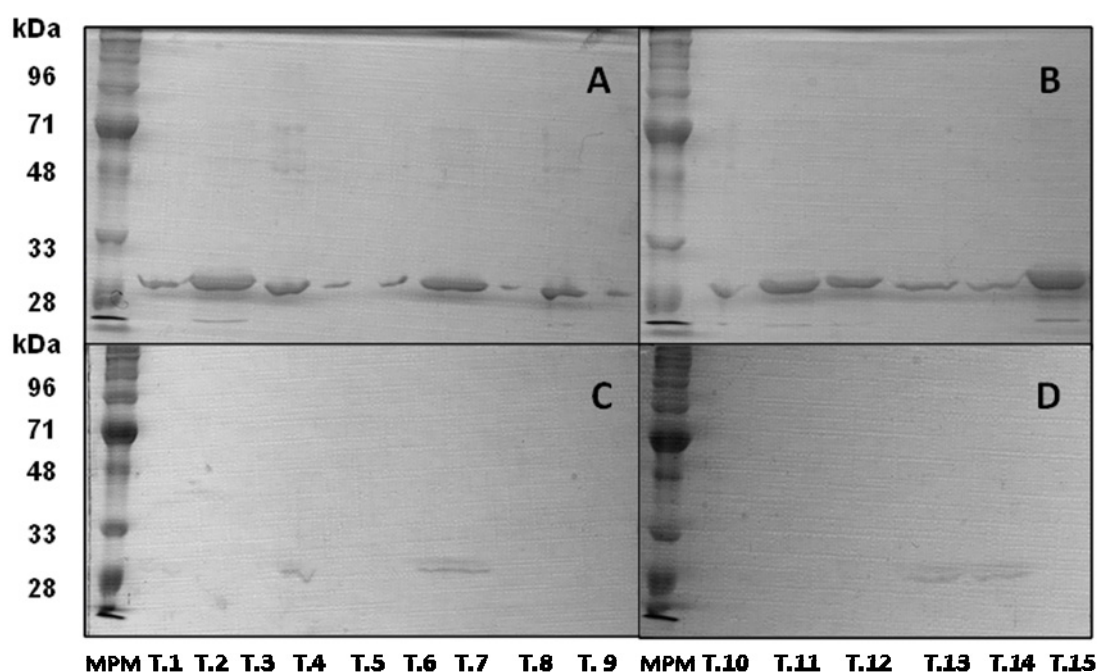
Además de visualizar en un gel de acrilamida los resultados de la desnaturalización para cada una de las concentraciones de urea, se cuantificó la cantidad de proteína obtenida para cada condición mediante  $Abs_{280}$  (tabla A1.1. del Anexo 1), obteniéndose aproximadamente la misma cantidad de proteína para los tampones de 8 M y 6 M urea y mucho más baja para 4 M, confirmando que la mejor concentración de urea a utilizar es 6 M. Cabe destacar que esta concentración de urea es elevada, lo que indica que la proteína aislada de los cuerpos de inclusión se encuentra muy mal plegada y que una vez desnaturalizada estará casi totalmente desplegada, lo que probablemente dificulte el proceso de renaturalización *in vitro*. Además, durante la realización de estas pruebas se observó que ROCO una vez desnaturalizado era muy inestable y agregaba con facilidad al diluir en pequeña medida el tampón de desnaturalización. También se pensó que su tamaño podría ser excesivo para su correcta renaturalización *in vitro*, por lo que las siguientes pruebas de replegado se realizaron únicamente con el dominio COR.

El replegado *in vitro* requiere que la proteína se encuentre lo más pura posible para evitar la formación de agregados proteicos, por lo que antes de la renaturalización se purifica la proteína mediante una cromatografía en columna o por FPLC utilizando una resina de afinidad de sefarosa-níquel, aprovechando la cola de 6 histidinas en el extremo N-terminal. En la Figura 3.3 se muestran los resultados obtenidos.



**Figura 3.3. Fracciones obtenidas mediante cromatografía de afinidad FPLC de las muestras de expresión del dominio COR en tampón urea 6 M.** Las 5 fracciones de eluidos se obtienen con concentración creciente de imidazol hasta 250 mM. Se observa que la mayor cantidad de dominio COR purificado se encuentra en la fracción del segundo eluido.

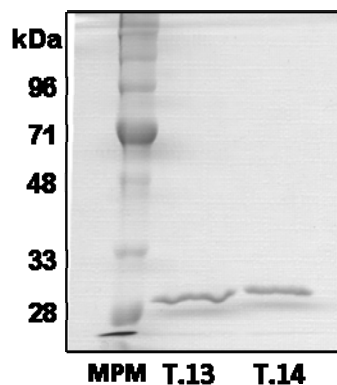
No obstante, durante este paso de purificación se perdió una cantidad significativa de proteína en la fracción no retenida y en los lavados, probablemente debido a que la urea dificulta la interacción con la resina. Para recuperar mayor cantidad de proteína se realizó otra cromatografía de la fracción no retenida pasándola por la columna varias veces antes de los lavados. Con la proteína purificada se realizó una pequeña prueba (1 mL) de replegado *in vitro* en distintos tampones, para determinar en cuál se obtiene mayor cantidad de proteína soluble. La composición de los distintos tampones probados se encuentra en el apartado 2.2.6 de Materiales y Métodos. El resultado de esta prueba se representa en la Figura 3.4.



**Figura 3.4. Pruebas del replegado del dominio COR en distintos tampones.** En los geles A y B se muestra la proteína presente en los sedimentos, mientras que las muestras de los geles C y D corresponden a los sobrenadantes. Se presentan ordenadas las muestras obtenidas con los tampones de reanturalización (tabla 2.2 de Materiales y Métodos) según se indica.

Como se observa en la figura, se obtiene proteína soluble con los tampones 1, 4, 7, 13 y 14, siendo los tres primeros de pH 6 y los dos últimos de pH 8,5. En un principio se realizó una prueba en un volumen mayor (10 mL) en los tampones 1, 4 y 7 ya que tienen un pH más distante al pI de la proteína, pero al cambiar el volumen de trabajo toda la proteína agregó sin quedar nada en la fracción soluble. Por este motivo se realizó el mismo procedimiento con los tampones 13 y 14, tras el replegado se

concentró la muestra mediante centricones (Figura 3.5). Pese a que los centricones permiten concentrar la proteína, también se pierde una cantidad significativa en este paso por agregación, que se observa a simple vista.



**Figura 3.5. Dominio COR renaturalizado presente en la fracción soluble concentrada del tampón 13 y 14 según se indica.**

Debido a que la cantidad de proteína obtenida en la fracción soluble para ambos tampones (13 y 14) es similar, pero que el tampón 13 es más sencillo en cuanto a su composición, se eligió este tampón para los posteriores replegados del dominio COR a mayor escala. Tras la renaturalización en 50 mL, se concentró la muestra y al mismo tiempo se sustituyó el tampón a uno más apropiado para los pasos posteriores (3 mL de Tris-HCl 100 mM a pH 8 y NaCl 0,5 M). Se cuantificó la concentración proteica de la muestra resultante mediante el método de Bradford (tabla A1.2. del Anexo 1), determinando que era de aproximadamente 20 mg/mL.

Curiosamente la renaturalización de COR se produjo cerca de su  $pI$  (8,6), lo que hace dudar de la predicción de este parámetro físico-químico o plantearse el hecho de que la proteína soluble obtenida no se halle en forma monomérica. Para comprobar el estado de agregación proteica, se realizó una cromatografía de exclusión molecular con la que se calculó que la  $M_m$  de la solución proteica resultado del replegado era aproximadamente de 235 kDa (el calibrado y el volumen de exclusión de la muestra se muestran en el Anexo 2), determinándose que la muestra estaba compuesta por oligómeros solubles formados por unas nueve unidades de COR. Tras esto se visualizó la proteína en un gel de acrilamida sin incluir  $\beta$ -mercaptoetanol (agente reductor) en el tampón de carga, pero no se observaron diferencias en el tamaño de la proteína con el correspondiente a su estado monomérico (27 kDa aproximadamente), por lo que se determinó que si bien la proteína estaba en forma de agregado soluble, éste no se

mantenía mediante puentes disulfuro. Por lo tanto, no se pudo determinar si estas asociaciones solubles de COR tenían connotaciones negativas para la estabilidad de su estructura dado que no estaban mantenidas por uniones covalentes.

### 3.1.4. Prueba de estabilidad proteica

Para determinar el tampón óptimo en que la proteína replegada es más estable y como control de la adecuada renaturalización de la proteína, se realizó una prueba de estabilidad (apartado 2.2.8 de Materiales y Métodos). Para determinar qué cantidad de proteína se debía usar en este ensayo, se realizó una prueba con diferentes diluciones de proteína (0,05, 0,1, 0,25, 0,5, 0,75 y 1 mg/mL) en tampón HEPES. Mediante el análisis de los resultados de este ensayo se observó que para todas aquellas concentraciones ensayadas había variación negativa de la fluorescencia según aumentaba la temperatura. Cabe destacar que se espera que el máximo de la curva sigmoide obtenida al representar la fluorescencia frente la temperatura, coincida con la temperatura a la que la proteína se encuentre totalmente desnaturalizada, pero en este caso se observó que este máximo se encontraba a temperatura ambiente, siendo menor la fluorescencia a temperaturas mayores como se observa en la Figura 3.6. Además de la representación gráfica, con los datos proporcionados por el *software* del instrumento se calcula para cada concentración la  $T_m$  (temperatura a la cual la mitad de la proteína se encuentra desnaturalizada) según la ecuación de Boltzmann, siendo en este caso valores negativos o inferiores a 20 °C.

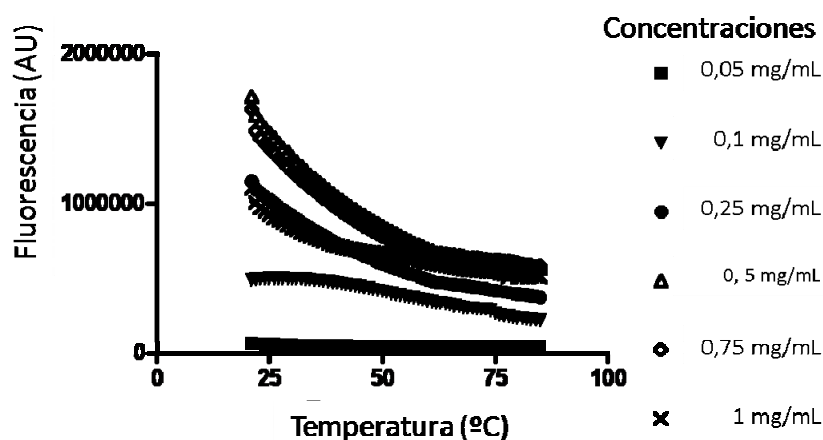


Figura 3.6. Representación gráfica de la fluorescencia frente a la temperatura en la prueba de estabilidad proteica.



Los resultados obtenidos podían explicarse de diversas formas: el ensayo no era útil porque el cromóforo no interaccionaba adecuadamente con la proteína de trabajo; la proteína se encontraba mal plegada y lo que se observaba en la gráfica era una pérdida de emisión respecto al tiempo; o en el medio estándar donde se realizó el ensayo la proteína era inestable. Para descartar esta última posibilidad, se realizó el ensayo de estabilidad en 48 tampones distintos (Anexo 3) utilizando la proteína a 0,25 mg/mL, por ser la concentración más baja que ofrece mayor respuesta, y añadiendo el cromóforo al final del proceso. Utilizando esta concentración los resultados fueron similares, sin encontrarse un comportamiento distinto para los diferentes tampones ensayados, por lo que se concluyó que o bien la proteína estaba mal plegada o que este método no era adecuado para su análisis. Con estos últimos resultados parece haberse comprobado que el replegado del dominio COR no ha dado buen resultado, ya que aunque sea soluble no parece ser estable. Sin embargo, se realiza un ensayo de *pull-down* con la muestra obtenida para intentar detectar la unión a otras proteínas.

### **3.1.5. Ensayo de *pull-down***

En primer lugar, se realizó un extracto proteico de cerebro de rata y se cuantificó mediante el método de Bradford, obteniéndose una concentración de aproximadamente 145 mg/mL de proteína (tabla A1.3 del Anexo 1). Utilizando este extracto y el dominio COR replegado se realizó el ensayo de *pull-down* en 10 mL de resina de sefarosa-níquel, tal y como se describe en el apartado 2.2.10 de Materiales y Métodos. De forma paralela se realizaron otras dos cromatografías de afinidad mediante el mismo procedimiento con la diferencia de que una no incluía el pase del extracto de cerebro de rata y la otra no incluía el pase del dominio COR, ambas se tomaron como controles negativos para descartar posibles artefactos en los resultados (posibles impurezas y productos de degradación de COR y proteínas del extracto de cerebro de rata que queden inespecíficamente unidas a la resina). Posteriormente, las fracciones de interés recogidas de cada cromatografía se concentraron mediante precipitación con acetona y se cargaron en un gel de acrilamida (Figura 3.7), con el fin de analizar mediante espectrometría de masas las bandas del gel que resultaran de interés.

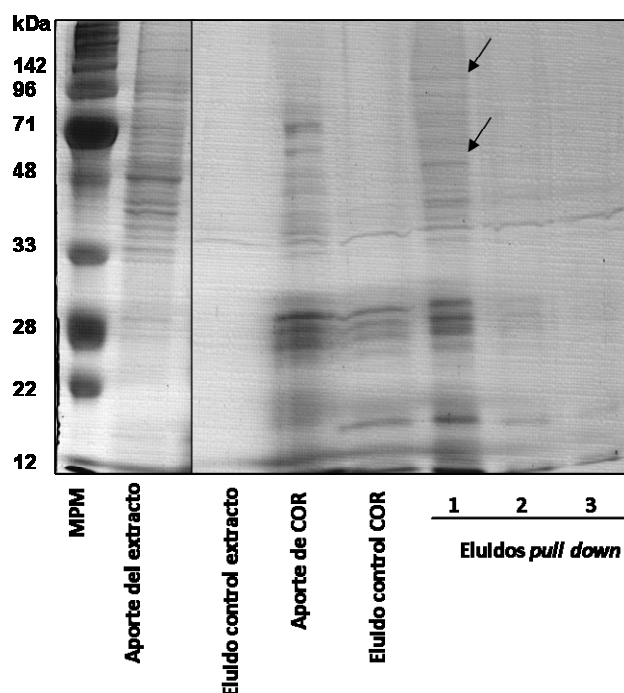


Figura 3.7. Ensayos de *Pull-down* utilizando dominio COR replegado *in vitro* y extracto de cerebro de rata. Las flechas señalan las bandas que podrían resultar de interés.

Como se observa en la figura anterior aunque hay alguna banda de interés que no está en el aporte de COR ni en los eluidos de los controles, estas proteínas se encuentran en muy baja cantidad para poder analizarlas por espectrometría de masas o secuenciarse por el N-terminal a partir de la extracción desde el gel.

### **3.2. Análisis *in silico* de la estructura terciaria de los dominios de LRRK2**

En paralelo al trabajo experimental con los distintos dominios de LRRK2, se analizó *in silico* (véase Materiales y Métodos apartado 2.2.12 y 2.3) la estructura secundaria y terciaria de cada dominio y el posible efecto que sobre ésta puedan tener las distintas mutaciones descritas para avanzar en el conocimiento sobre esta proteína asociada a la EP. Como se ha comentado en la introducción, la definición de los distintos dominios de LRRK2 es controvertida, por lo que para el análisis *in silico* de la estructura de estos dominios y de sus mutaciones se seleccionó la definición de dominio más adecuada en cada caso en cuanto a estructura tridimensional obtenida utilizando los resultados de la predicción de estructura terciaria. De este modo se decidió utilizar las definiciones de dominios que aparecen en la Figura 3.8, que coinciden con los seleccionados en la Figura 1.1 de la Introducción. Cabe destacar que para el dominio ROC se utilizó finalmente la misma definición que la utilizada para su elucidación estructural, por considerarse la más adecuada.

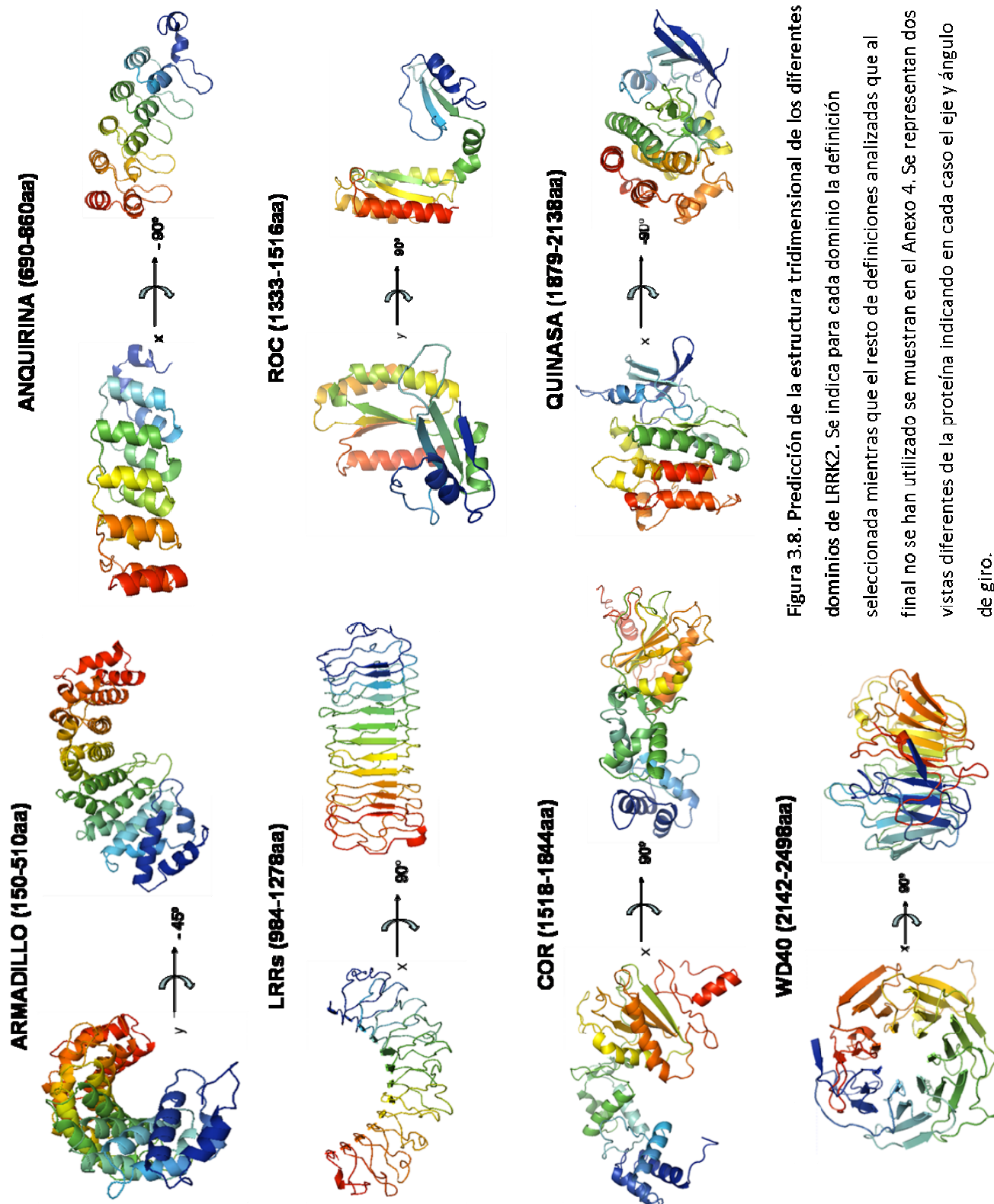


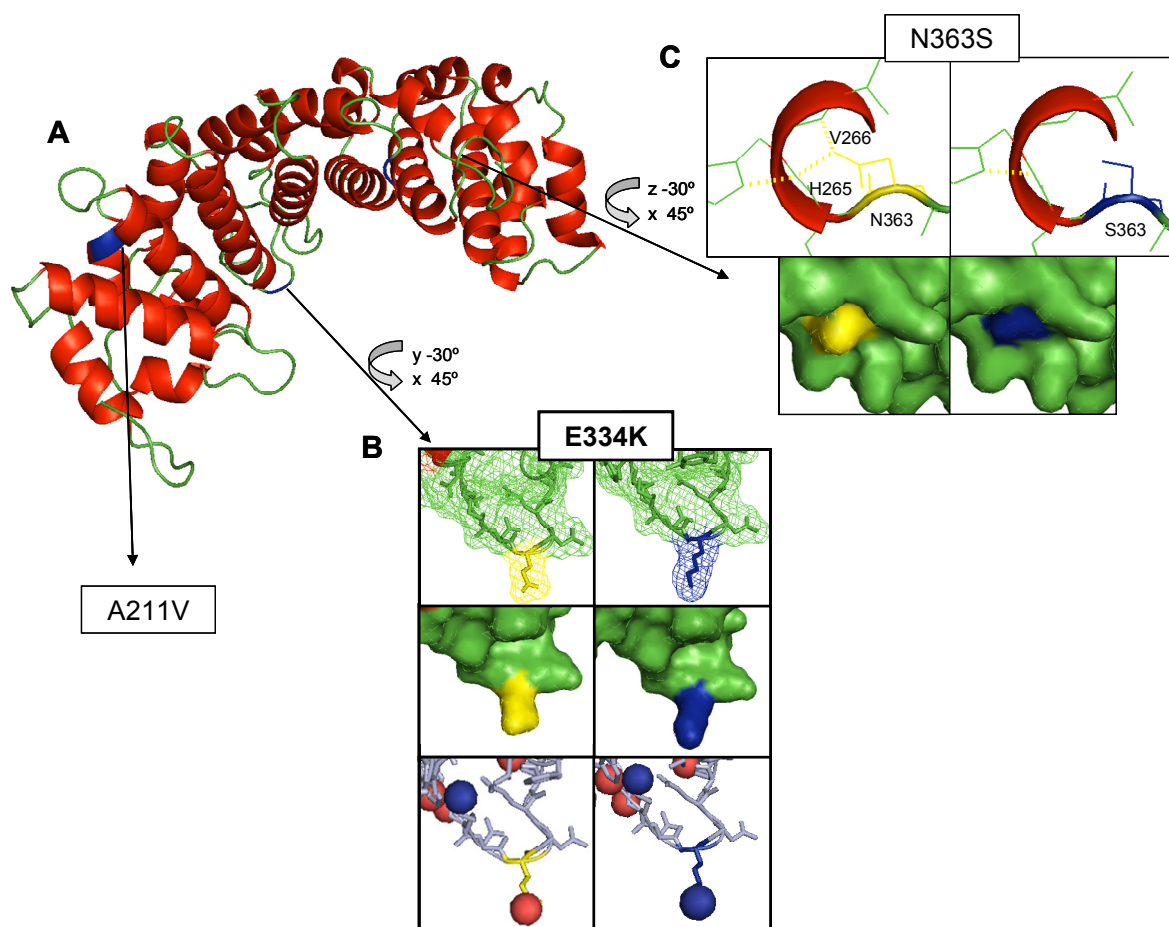
Figura 3.8. Predicción de la estructura tridimensional de los diferentes dominios de LRRK2. Se indica para cada dominio la definición seleccionada mientras que el resto de definiciones analizadas que al final no se han utilizado se muestran en el Anexo 4. Se representan dos vistas diferentes de la proteína indicando en cada caso el eje y ángulo de giro.

### 3.2.1. Análisis del dominio armadillo

Según la predicción de la estructura secundaria (Figura A5.1 del Anexo 5) y terciaria (Figura 3.9-A) este dominio está formado por 21 hélices- $\alpha$  agrupadas de tres en tres formando superhélices y unidas entre sí mediante lazos, de forma que el conjunto adquiere una estructura curva y los lazos quedan más accesibles al solvente. La predicción de la estructura terciaria se realizó mediante *threading* (programa *HH search* v.1.5b) utilizando como molde la importuna- $\alpha$  de *Saccharomyces cerevisiae* (código PDB 1wa5B).

Para este dominio se han descrito tres cambios de aminoácido: A211V, E334K y N363S; de los cuales E334K es una mutación patogénica. Tras el estudio del posible efecto que estas sustituciones puedan tener en la estructura del dominio se ha determinado que los cambios E334K y N363S afectan a la superficie del dominio (Figura 3.9-B y C, respectivamente), siendo en principio más significativa la E334K que determina además un cambio de carga de negativo a positivo (Figura 3.9-B) y la N363S que determina la pérdida de dos puentes de hidrógeno establecidos por el residuo N363 con los aminoácidos H265 y V266 (Figura 3.9-C). Estos dos residuos forman parte de una hélice- $\alpha$  cercana a un lazo, lo que podría repercutir en la rigidez de la posición del lazo respecto la hélice.

Cabe destacar que aunque dos de los cambios descritos para este dominio (E334K y N363S) provocan variaciones en la estructura, sólo E334K es patogénico, quizá debido a que la posición 334 se encuentra más accesible al solvente que la 363 (no está tan próxima a una hélice- $\alpha$ ), y por tanto el cambio provocado en la superficie es más importante. Además un cambio de carga en la superficie de un dominio de interacción, como es el caso de armadillo, puede resultar muy determinante en cuanto a las moléculas con las cuales puede interaccionar.

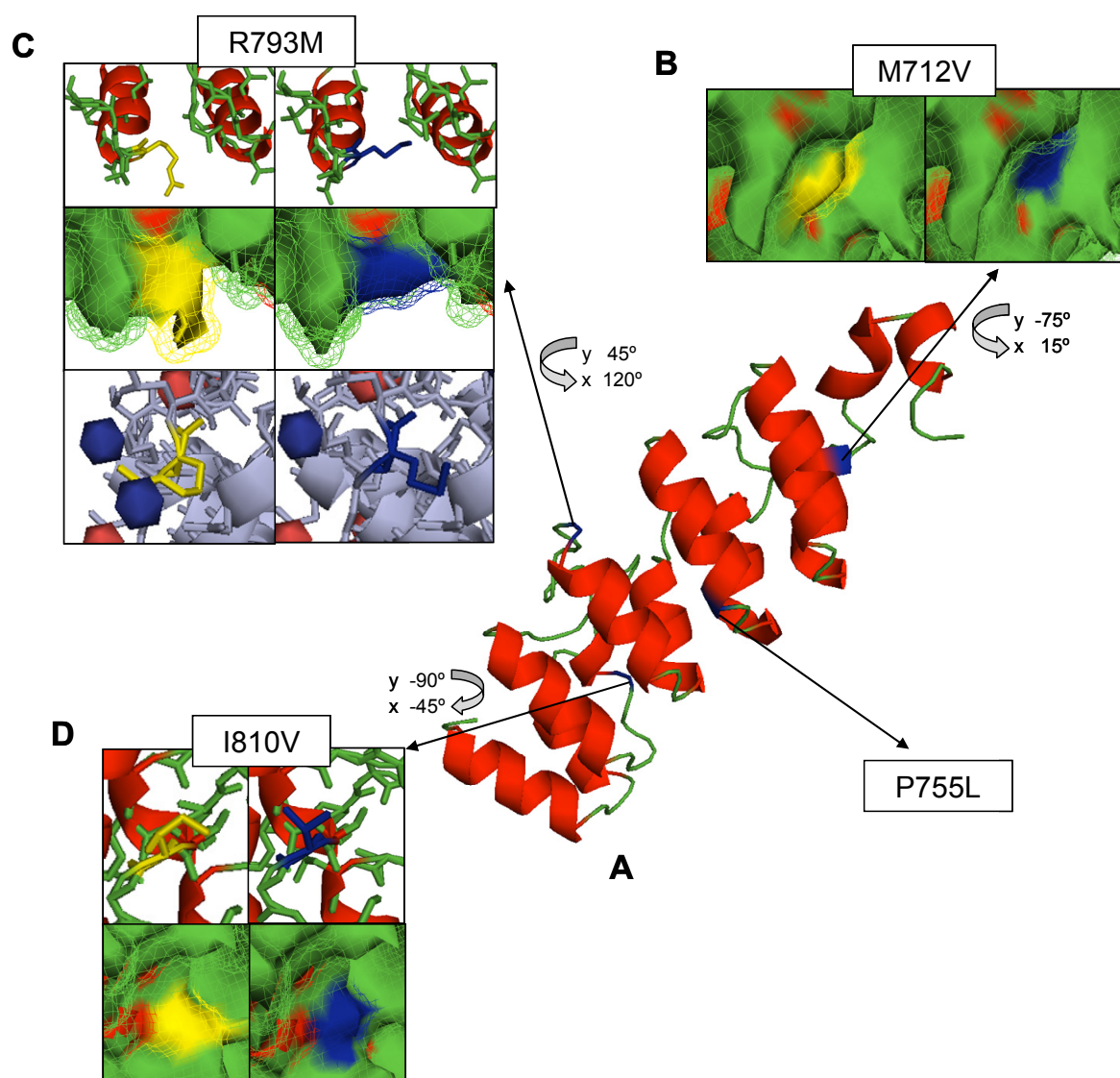


**Figura 3.9. Predicción de la estructura terciaria del dominio armadillo de LRRK2 mediante *threading* y representación del efecto de los cambios descritos.** Las hélices- $\alpha$  se representan en rojo y los lazos en verde. El aminoácido silvestre se representa en amarillo y el mutante en azul. Los puentes de hidrógeno se muestran como líneas amarillas discontinuas, mientras que las cargas positivas y negativas se representan como esferas azules y rojas, respectivamente. Se indican en negrita las mutaciones patológicas para distinguirlas de los polimorfismos.

### 3.2.2. Análisis del domino anquirina

Según la predicción de la estructura secundaria (Figura A5.2 del Anexo 5) y terciaria (Figura 3.10-A), este dominio está formado por 11 hélices- $\alpha$  unidas mediante lazos, de forma que los lazos quedan más accesibles al solvente. La predicción de la estructura terciaria se realizó mediante *threading* (programa SPARKS2) utilizando como molde la ganquirina (código PDB 1uohA).

Para este dominio se han descrito cuatro polimorfismos: M712V, P755L, R793M y I810V. Tras el estudio de estos cambios se ha determinado que M712V, R793M y I810V afectan a la superficie del dominio (Figura 3.10-B, C y D, respectivamente), siendo este efecto más importante para R793M, que además determina la pérdida de una carga positiva en la superficie del dominio (Figura 3.10-C). No obstante, estos cambios no se han asociado a la patogénesis de la EP (Tomiya *et al.*, 2008).



**Figura 3.10. Predicción de la estructura terciaria del dominio anquirina de LRRK2 mediante *threading* y representación del efecto de los cambios descritos.** Las hélices- $\alpha$  se representan en rojo y los lazos en verde. El aminoácido silvestre se representa en amarillo y el mutante en azul. Las cargas positivas y negativas se representan como esferas azules y rojas, respectivamente.

### 3.2.3. Análisis del dominio LRR

La predicción de la estructura secundaria en este caso parece menos correcta, ya que las hélices- $\alpha$  y láminas- $\beta$  son pocas y muy cortas (Figura A5.3 del Anexo 5). La predicción de la estructura terciaria se realizó por homología de secuencia utilizando como molde Lingo1 (código PDB 2id5C). La predicción (Figura 3.11-A) muestra que el dominio está formado por 13 láminas- $\beta$  paralelas unidas entre sí por extensos lazos y que en conjunto adquiere una estructura curva con elevada accesibilidad al solvente.

Se han descrito ocho sustituciones de aminoácido a lo largo de este dominio, de las cuales cuatro se consideran patogénicas (Q1111H, I1122V, I1192V y S1228T). De los cambios descritos R1067Q, S1096C, Q1111H y A1151T producen cambios en la superficie (Figura 3.11-B, C, D y E, respectivamente), y R1067Q implica además la pérdida de una carga positiva (Figura 3.11-B), mientras que A1151T provoca la aparición de dos puentes de hidrógeno con los aminoácidos L1125 y E1150 (Figura 3.11-E), perteneciendo el primero a un lazo próximo, lo que podría implicar una mayor rigidez en la conexión de ambos lazos. La sustitución S1228T también afecta a la composición de puentes de hidrógeno (Figura 3.11-G), de forma que S1228 forma un puente de hidrógeno con S1203, N1230 y S1250 mientras que T1228 forma dos puentes de hidrógeno con S1203 y uno con N1230, lo que podría fortalecer la interacción del lazo que contiene el cambio con el que contiene el aminoácido S1203 y a su vez debilitarla con el que contiene el aminoácido S1250. Por último, el cambio H1216R implica la ganancia de una carga positiva en la superficie del dominio (Figura 3.11-F). Cabe destacar que todos los polimorfismos descritos parecen afectar a la estructura del dominio, sin embargo de las cuatro mutaciones patogénicas sólo se han observado efectos estructurales en dos (Q1111H y S1228T). El hecho de no encontrar efectos estructurales para otras mutaciones puede deberse a que las posiciones de estas sustituciones estén implicadas en las interacciones con otras proteínas o en la dimerización de LRRK2, por lo que no requerirían representar importantes efectos sobre la estructura del dominio. Por ejemplo, en el caso de I1122V y I1192V, aunque la sustitución del aminoácido no cambia el estado de carga o la hidrofobicidad, la diferencia de tamaño de la cadena lateral podría ser suficiente para generar cambios en las interacciones con otras proteínas.



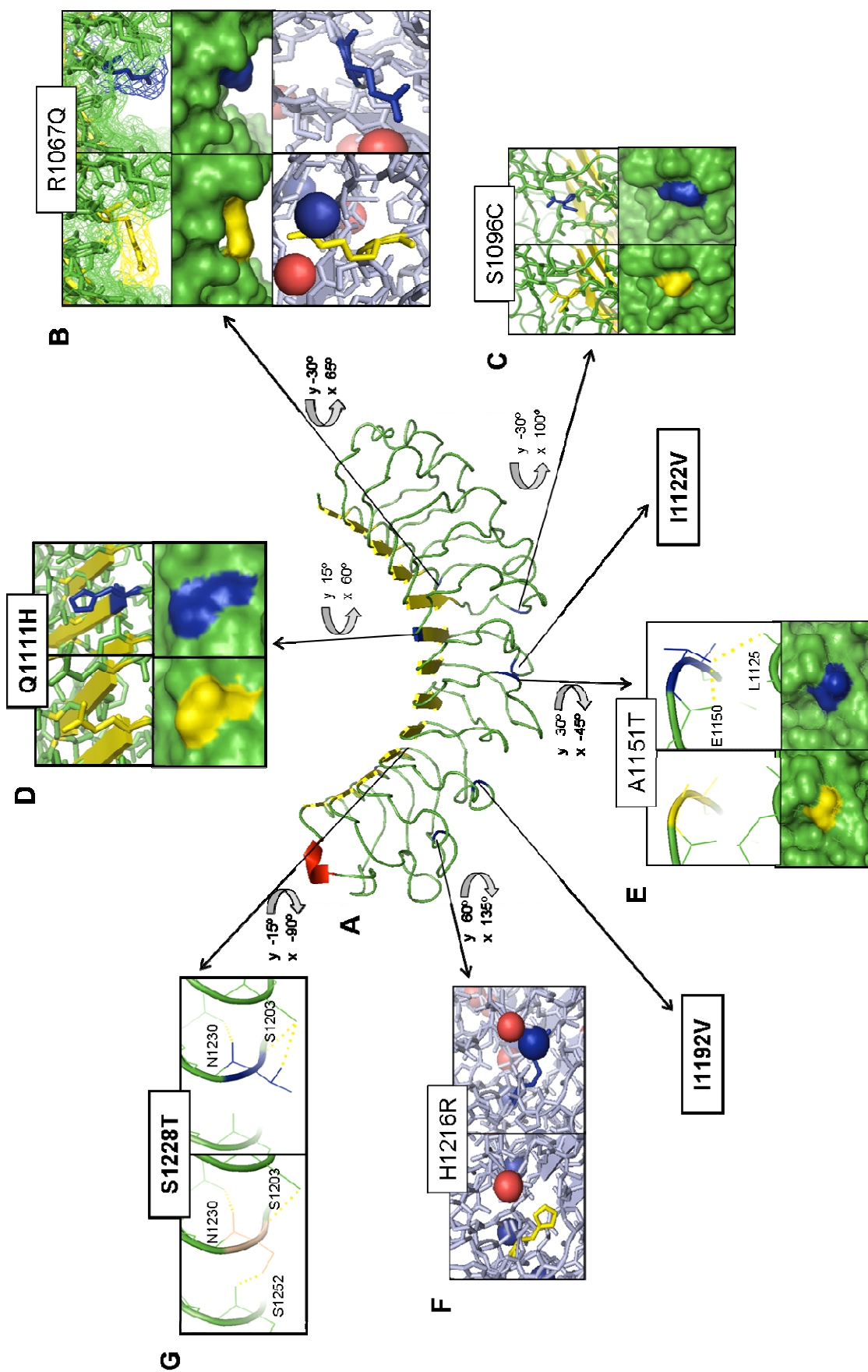


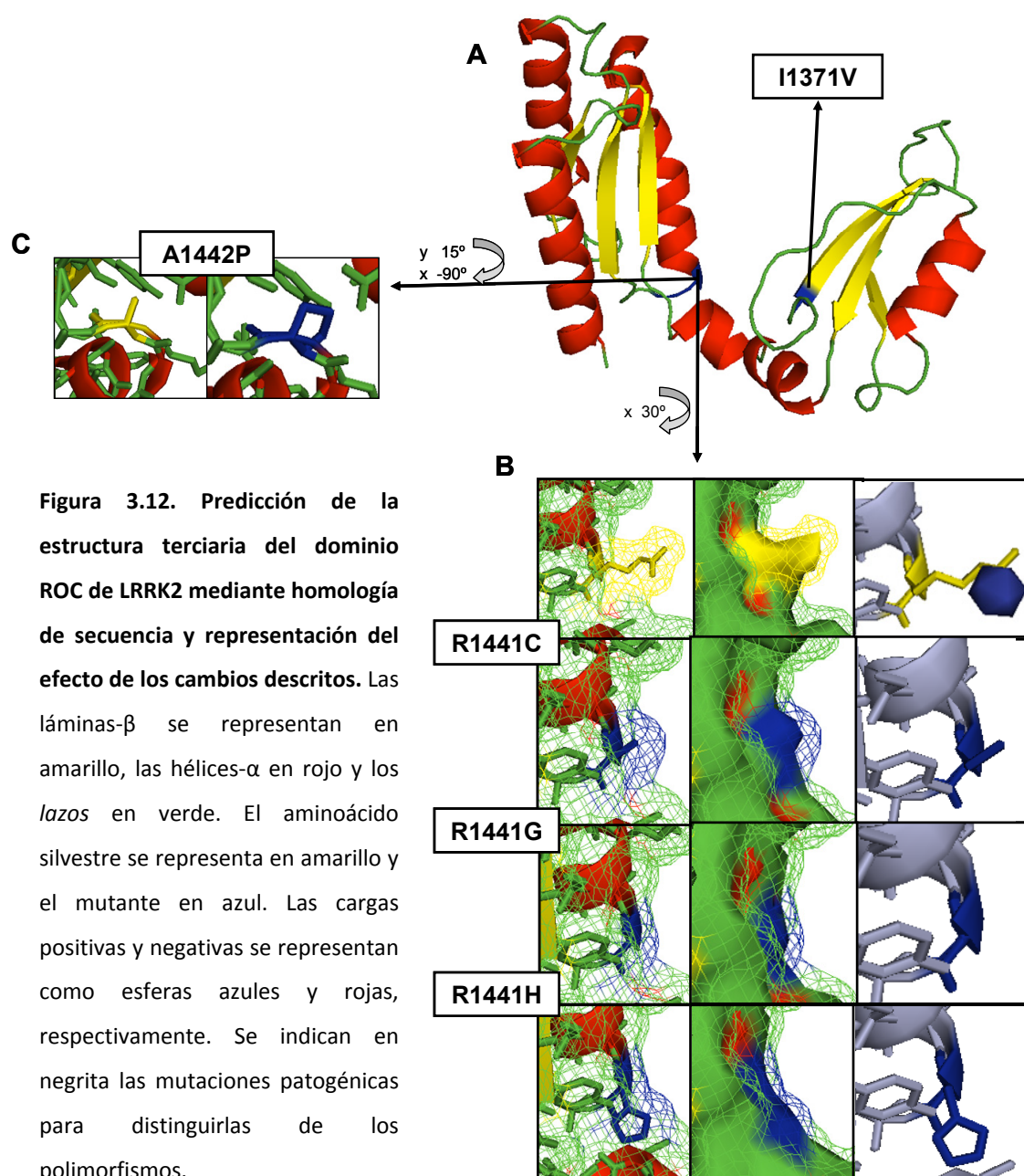
Figura 3.1.1. Predicción de la estructura terciaria del dominio LRR de LRRK2 mediante homología de secuencia y representación del efecto de los cambios descritos. Las láminas- $\beta$  se representan en amarillo, las hélices- $\alpha$  en rojo y los lazos en verde. El aminoácido silvestre se representa en amarillo y el mutante en azul. Los puentes de hidrógeno se muestran como líneas amarillas discontinuas, mientras que las cargas positivas y negativas se representan como esferas azules y rojas, respectivamente. Se indican en negrita las mutaciones patogénicas para distinguirlas de los polimorfismos.



### 3.2.4. Análisis del dominio ROC

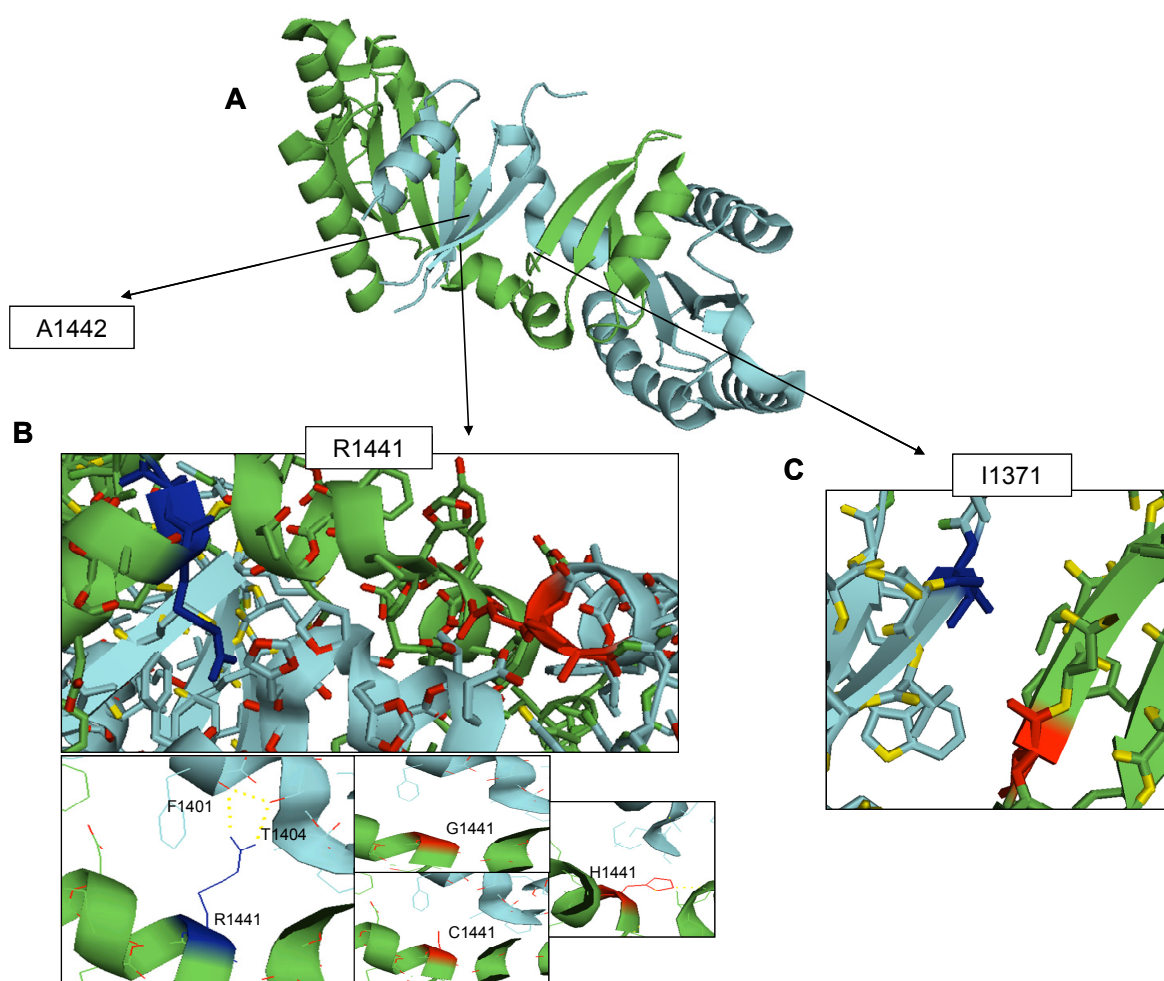
Como se ha comentado en la introducción, la estructura de este dominio ha sido resuelta mediante cristalografía de rayos X (Deng *et al.*, 2008), determinándose que está formado por seis láminas- $\beta$  y cinco hélices- $\alpha$ . Aun así, se realizaron y analizaron las distintas predicciones de estructura de la forma silvestre y las sustituciones de aminoácidos descritas para ROC. En la predicción de la estructura secundaria y terciaria se observan cinco hélices- $\alpha$  y seis láminas- $\beta$  unidas mediante lazos, encontrándose estos últimos más accesibles al solvente (Figura A5.4 del Anexo 5 y la Figura 3.12-A, respectivamente). Así pues, la estructura predicha realizada por homología de secuencia, utilizando como molde la propia estructura del dominio de la proteína de humanos que ha sido resuelta (código PDB 2zej), se corresponde con la resuelta por difracción.

Para este dominio se han descrito cinco cambios asociados con la patogénesis de la EP, de los cuales tres recaen sobre un mismo aminoácido. Como se observa en la Figura 3.12-B los tres cambios descritos para la posición 1441 de LRRK2 afectan a la conformación de la superficie del dominio ROC e implican la pérdida de una carga positiva. También podría dificultar la dimerización de la proteína, ya que esta posición se encuentra en la interfase de la interacción. Como se muestra en la Figura 3.13-B, R1441 forma puentes de hidrógeno con aminoácidos del otro monómero (aminoácidos F1401 y T1404) cosa que no ocurre en las sustituciones R1441G/C/H; de hecho, la mutación R1441H implica la ganancia de un puente de hidrógeno dentro del propio monómero (la disposición de puentes de hidrógeno para la posición 1441 ha sido descrita por Vasanti, *et al.* 2009).



Recientemente, se han identificado algunos sitios de autofosforilación de LRRK2, siendo uno de ellos el aminoácido T1404 (que forma puentes de hidrógeno con R1441). Kamikawaji *et al.* (2009) postulan que la fosforilación de la T1404 es requerida para la fosforilación del residuo T1491, involucrada en la unión a GTP y por lo tanto en la activación de la quinasa. A partir de esta información se podría hipotetizar que la posición T1404 debe quedar más accesible a la fosforilación al perder las interacciones con R1441, y por tanto las sustituciones en la posición R1441 podrían implicar el incremento de la actividad de la quinasa a través de este mecanismo. Por otra parte,

aunque la sustitución I1371V no presentó ningún efecto aparente en la estructura de la proteína, esta posición también se encuentra en la interfase de la interacción entre los dos monómeros (Figura 3.13-A y B) pudiendo afectar a la misma. Por su parte la sustitución A1442P afecta a la conformación de la hélice- $\alpha$  en la que se encuentra al introducir un aminoácido cíclico en su estructura, lo que también podría afectar a la dimerización del dominio.



**Figura 3.13. Representación de la estructura dimérica del dominio ROC resuelta por difracción de rayos X (Deng *et al.*, 2008).** Se indica cada monómero con un color (verde, azul) y se posicionan las mutaciones descritas en el monómero en verde, ampliando la región de los residuos R1441 y I1371 de ambas cadenas, ya que estos aminoácidos se localizan en la interfase de unión de ambos monómeros. En la región 1441 se analizan los puentes de hidrógeno (líneas amarillas discontinuas) tanto con el aminoácido silvestre como en los mutantes descritos.

Cabe destacar que lo discutido anteriormente ha sido fruto del análisis realizado en este trabajo. Sin embargo, al haberse determinado la estructura real de este dominio los estudios sobre el efecto de las mutaciones asociadas a la EP han sido muy

extensos. Así pues, se ha descrito que el efecto en la estructura causado por los cambios R1441C/G/H y A1442P podría dirigir un desplazamiento de los residuos R1398 y D1394 importantes en la interacción del fosfato- $\gamma$  con el GTP y el ion  $Mg^{+2}$  y por tanto en la actividad GTPasa del dominio, lo que justificaría la disminución de esta actividad e incremento de la quinasa para los mutantes descritos; y que el cambio I1371V también determina la pérdida de interacciones hidrofóbicas importantes en la interfase del dímero (Deng *et al.*, 2008).

El hecho de que todas las mutaciones descritas para este dominio se hayan asociado con la EP destaca la importancia de esta GTPasa en la función de la proteína, ya sea porque afecta a su actividad y/o a la unión de sustratos, o porque afecta a la dimerización de la proteína.

### 3.2.5. Análisis del dominio COR

Según la predicción de la estructura secundaria (Figura A5.5 del Anexo 5) y terciaria (Figura 3.14-A) este dominio está formado por nueve hélices- $\alpha$  y seis láminas- $\beta$  unidas por lazos, siendo éstos más accesibles al solvente. La mitad N-terminal del dominio contiene seis de las hélices- $\alpha$  y dos pequeñas láminas- $\beta$ , mientras que la región C-terminal tiene tres hélices- $\alpha$  y cuatro láminas- $\beta$  antiparalelas consecutivas. La predicción de la estructura terciaria se realizó mediante *threading* (programa *HH search* v1.5b) utilizando como molde la estructura resuelta por Gotthardt *et al.* (2008) del dominio homólogo en *Chlorobium tepidum* (código PDB 3dpuA). Respecto a la estructura descrita para el molde, la predicción del dominio COR de LRRK2 difiere principalmente en su mitad C-terminal, que es la zona del dominio por la cual la proteína de *C. tepidum* dimerizaba, esto apoya a los datos estructurales que demuestran que la dimerización de LRRK2 se da a través de su dominio ROC-GTPasa.

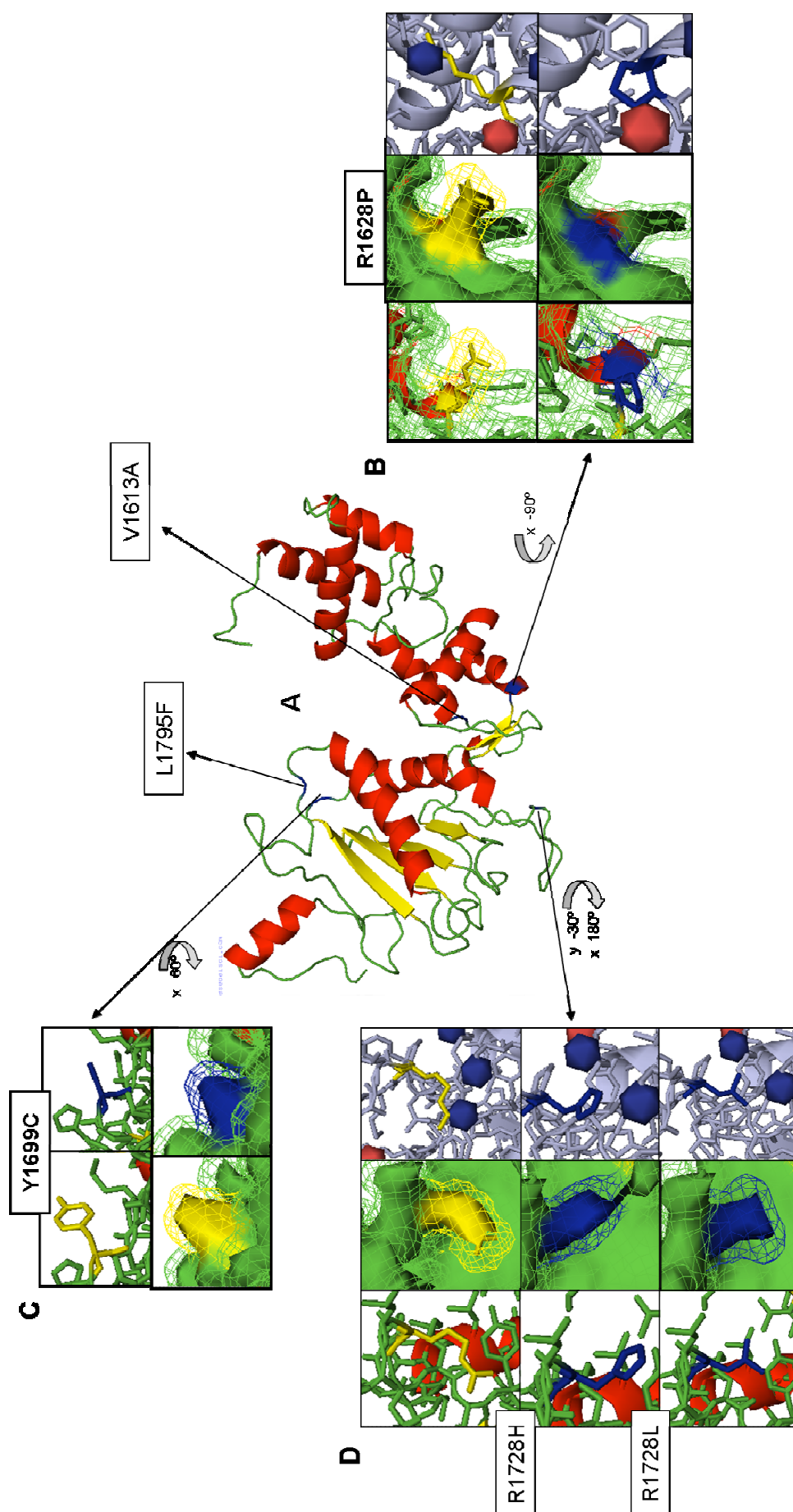


Figura 3.14. Predicción de la estructura terciaria del dominio COR de LRRK2 mediante *threading* y representación del efecto de los cambios descritos. Las láminas- $\beta$  se representan en amarillo, las hélices- $\alpha$  en rojo y los lazos en verde. El aminoácido silvestre se representa en azul, y las cargas positivas y negativas se muestran como esferas azules y rojas, respectivamente. Se indican en **negrita** las mutaciones patogénicas para distinguir de los polimorfismos.

Para este dominio se han descrito dos mutaciones asociadas con la EP: R1628P y Y1699C, y cuatro polimorfismos; dos de los cuales recaen sobre la misma posición. Tras el estudio de estos cambios se ha determinado que las sustituciones R1628P, Y1699C y R1728H/L afectan a la superficie de la proteína (Figura 3.14-B, C y D, respectivamente), y R1628P y R1728H/L también implican la pérdida de una carga positiva (Figura 3.14-B y D, respectivamente). No se han observado efectos en el número de puentes de hidrógeno (datos no mostrados).

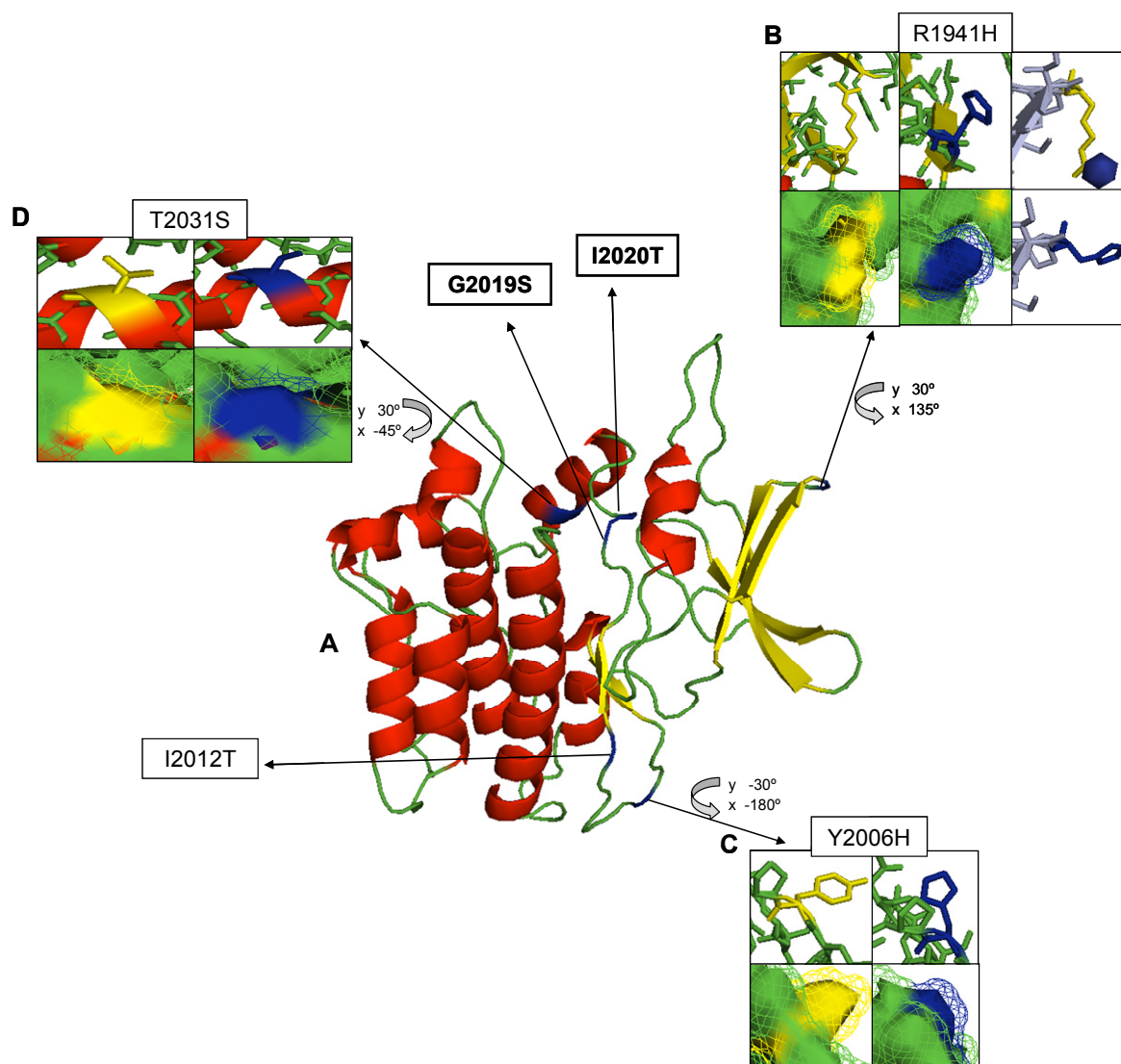
Las mutaciones patogénicas descritas afectan a la estructura predicha para el dominio, así como los polimorfismos R1728H/L. El motivo por el cual sólo R1628P y Y1699C se han asociado a la EP podría ser que están en la interfase de unión con el dominio ROC, y por tanto podrían afectar a la transducción de la señal entre la GTPasa y la quinasa, como ya ha sido postulado por Gotthardt *et al.* (2008).

#### **2.2.6. Análisis del dominio quinasa**

Según la predicción de la estructura secundaria (Figura A5.6 del Anexo 5) y terciaria (Figura 3.15-A), este dominio está formado por siete hélices- $\alpha$  y por seis láminas- $\beta$  unidas a través de lazos, que quedan más accesibles al solvente. La predicción de estructura terciaria se realizó mediante homología de secuencia utilizando como molde la tirosina quinasa humana ZAP-70 (código PDB 2ozoA).

Se han descrito seis mutaciones en este dominio catalítico, de las cuales dos son patogénicas (G2019S y I2020T). Tras el análisis del efecto de estas sustituciones se ha determinado que R1941H, Y2006H y T2031S implican una variación en la superficie del dominio (Figura 3.15-B, C y D, respectivamente), y R1941H además implica la pérdida de una carga positiva (Figura 3.15-B).





**Figura 3.15. Predicción de la estructura terciaria del dominio quinasa de LRRK2 mediante homología de secuencia y representación del efecto de los cambios descritos.** Las láminas- $\beta$  se representan en amarillo, las hélices- $\alpha$  en rojo y los lazos en verde. El aminoácido silvestre se representa en amarillo y el mutante en azul. Las cargas positivas y negativas se representan como esferas azules y rojas, respectivamente. Se indican en **negrita** las mutaciones patogénicas para distinguirlas de los polimorfismos.

No se han observado cambios en la predicción de estructura para las mutaciones patogénicas, aunque cabe destacar que estas sustituciones son las que se encuentran en el lazo de activación. G2019S está en la región conservada DFG/DYG, por lo que cambios en la cadena lateral de estas posiciones aunque no representen un efecto en la estructura del dominio pueden afectar a su activación. De hecho se ha postulado que las sustituciones G2019S y I2020T dirigen un cambio conformacional que mimetiza el estado activo de la quinasa, siendo este cambio mayor para G2019S, ya que

entorpece la flexibilidad característica de la región conservada en la que se encuentra al sustituir una de las glicinas. Esto justificaría que el aumento de la actividad quinasa sea más pronunciado para esta mutación que para I2020T, además G2019S genera un sitio adicional de autofosforilación que dirige también a la activación del dominio (Greggio y Cookson, 2009; Vasanti *et al.*, 2009). Cabe destacar que en algunos trabajos en los que también se ha realizado el modelado del dominio quinasa de LRRK2 se han reportado cambios estructurales en el lazo de activación debidos a las mutaciones patogénicas (Marín, 2006), si bien es cierto que para el modelado de cada una de las versiones del dominio se utilizaron moldes distintos.

### 3.2.7. Análisis estructural del dominio WD40

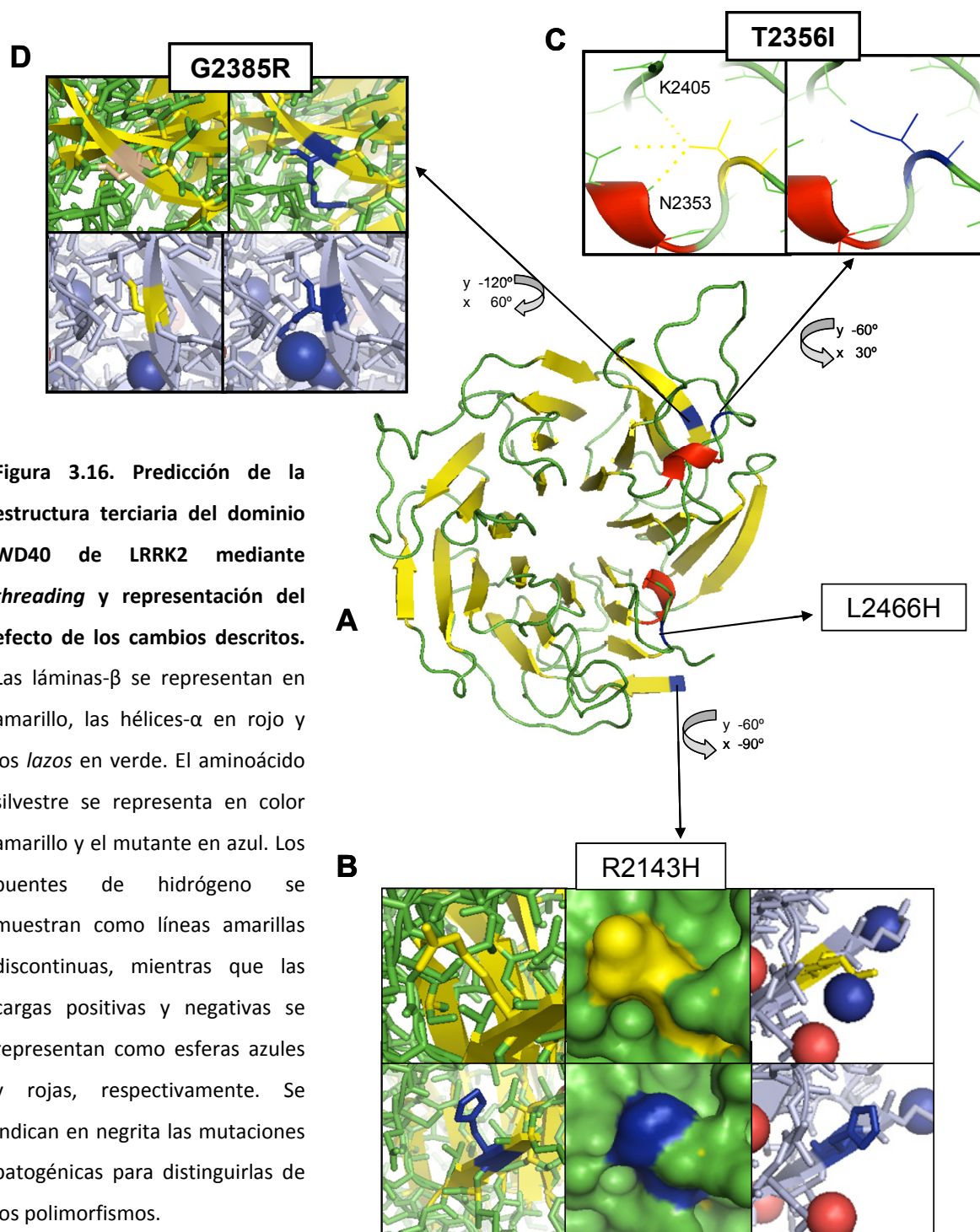
Según la predicción de la estructura secundaria (Figura A5.7 del Anexo 5) y terciaria (Figura 3.16) este dominio está formado por láminas- $\beta$  unidas mediante lazos, de forma que los lazos quedan accesibles al solvente. Como ya ha sido descrito (Mata *et al.*, 2006), se observa que este dominio adquiere una estructura  *$\beta$ -propeller* (siete repeticiones dispuestas en círculo a través de un eje central, donde cada repetición se compone por cuatro láminas- $\beta$  antiparalelas conectadas mediante lazos). La predicción de la estructura terciaria se realizó mediante *threading* (programa SP3) utilizando como molde la holoenzima paf-ah de ratón (código PDB 1vyhC).

Para este dominio se han descrito cuatro cambios de aminoácido dos polimorfismos (R2143H y L2466H) y dos mutaciones patogénicas (T2356I y G2385R). Tras el estudio del efecto de estas sustituciones se ha determinado que el polimorfismo R2143H afecta tanto a la superficie del dominio como a la distribución de cargas (Figura 3.16-B), ya que implica la pérdida de una carga negativa. La mutación T2356I implica la pérdida de tres puentes de hidrógeno, uno con el aminoácido K2405 y dos con N2353 (Figura 3.16-C), lo que podría causar un aumento de la flexibilidad al lazo donde se encuentra. Por último la mutación G2385R podría afectar a la interacción con los residuos cercanos, dado el aumento del tamaño de la cadena lateral del aminoácido, además de la ganancia de una carga positiva (Figura 3.16-D).

En este caso se observa la mayor importancia de las mutaciones patogénicas, ya que varían propiedades como la carga (G2385R) o la disposición de interacciones



débiles (T2356I). Además, tridimensionalmente se localizan cerca la una de la otra, indicando que su posición podría relacionarse con la función del dominio. El polimorfismo R2143H también parece tener un efecto en la estructura del dominio, pero hay que tener en cuenta que aunque cause la pérdida de una carga positiva, se sustituye un aminoácido ácido por otro, por lo que no cambian esencialmente las propiedades de la posición 2143 con la sustitución.



La siguiente tabla muestra un resumen de los datos obtenidos a partir del análisis *in silico* de la estructura de cada uno de estos dominios, indicando la importancia de la sustitución, la localización y el efecto en la estructura observado para las mutaciones descritas de cada dominio. La importancia del cambio hace referencia a lo significativo que resulta el cambio en cuanto a las propiedades físico-químicas (hidrofobicidad, carga, diferencias estéricas...) de los aminoácidos sustituidos. Para determinar lo significativa que puede ser la sustitución de un aminoácido por otro, se ha utilizado la matriz de Blosum 62, considerando que los cambios con valores cero o positivos tienen poca importancia mientras que los valores negativos se consideran de importancia media o alta.

**Tabla 3.3. Análisis de las sustituciones de aminoácidos en los distintos dominios de LRRK2.**

Dominio	Mutaciones	Importancia del cambio	Localización	Efecto en la estructura
ARMADILLO	A211V	Baja	hélice- $\alpha$	No se observa
	E334K	Alta (ácido por básico)	lazo	Cambio en la superficie y en la carga
	N363S	Baja	lazo cercano a una hélice- $\alpha$	Cambio en la superficie y en los puentes de H
ANQUIRINA	M712V	Baja	final hélice- $\alpha$	Cambio en la superficie
	P755L	Alta (cíclico por alifático)	final hélice- $\alpha$	No se observa
	R793M	Media (básico por azufrado)	lazo	Cambio en la superficie y en la carga
	I810V	Baja	lazo cercano a hélice- $\alpha$	Cambio en la superficie
LRRs	R1067Q	Alta (básico por ácido)	lazo	Cambio en la superficie y en la carga
	S1096C	Alta (polar por azufrado)	lazo	Cambio en la superficie
	Q1111H	Baja	lámina- $\beta$	Cambio en la superficie
	I1122V	Baja	lazo	No se observa
LRRs	A1151T	Baja	lazo	Cambio en la superficie y en los puentes de H

Dominio	Mutaciones	Importancia del cambio	Localización	Efecto en la estructura
	I1192V	Baja	lazo	No se observa
	H1216R	Baja	lazo	Cambio en la carga
	S1228T	Baja	lazo	Cambio en los puentes de H
ROC	I1371V	Baja	final lámina- $\beta$	No se observa
	R1441C	Media (ácido por azufrado)	final hélice- $\alpha$	Cambio en la superficie, en la carga y en los puentes de H $\rightarrow$ efecto en la dimerización
	R1441G	Media (ácido por apolar)		
	R1441H	Baja		
	A1442P	Media (apolar por cíclico)	lazo cercano a hélice- $\alpha$	Posible cambio en la hélice- $\alpha$ y efecto en la dimerización.
COR	V1613A	Baja	lazo cerca de hélice- $\alpha$	No se observa
	R1628P	Media (ácido por cíclico)	final hélice- $\alpha$	Cambio en la superficie y en la carga.
	Y1699C	Media (aromático por azufrado)	lazo	Cambio en la superficie.
	R1728H	Baja	lazo	Cambio en la superficie y en la carga.
	R1728L	Media (ácido por apolar)		
COR	L1795F	Baja	lazo	No se observa

Dominio	Mutaciones	Importancia del cambio	Localización	Efecto en la estructura
QUINASA	R1941H	Baja	lazo cercano a lámina- $\beta$	Cambio en la superficie y en la carga
	Y2006H	Baja	lazo (catalítico)	Cambio en la superficie
	I2012T	Baja	lazo (catalítico) cercano a lámina- $\beta$	No se observa
	G2019S	Baja	lazo de activación	No se observa
	I2020T	Baja	lazo de activación	No se observa
	T2031S	Baja	hélice- $\alpha$	Cambio en la superficie
WD40	R2143H	Baja	lámina- $\beta$	Cambio en la superficie y en la carga
	T2356I	Media (polar por apolar)	lazo	Cambio en los puentes de H
	G2385R	Media (apolar por ácido)	lámina- $\beta$	Cambio en la carga
	L2466H	Media (apolar por ácido)	lazo	No se observa

Las mutaciones patogénicas se muestran sombreadas para distinguirlas de los polimorfismos.

Muchas de las mutaciones descritas implican la pérdida o la ganancia de un residuo arginina, conllevando la pérdida o la ganancia de una carga positiva. A su vez para los dominios LRR y WD40 se ha descrito la presencia de residuos de arginina conservados en su superficie implicados en su función de interacción con otras proteínas (Mata *et al.*, 2006), por lo que estas mutaciones podrían estar implicadas en cambios en la especificidad de unión a distintas proteínas.

Por otra parte, aunque se han descrito varios polimorfismos a lo largo de LRRK2 no todos ellos se han asociado a mutaciones patogénicas, pero en el análisis *in silico* muchos de estos han mostrado tener efecto sobre la estructura de la proteína. Esto se podría explicar si estos polimorfismos se encuentran en regiones de la proteína que no estén implicadas en su función como la unión a otros sustratos, la actividad catalítica o la dimerización. También es posible que el efecto de estos polimorfismos fuera compensado por una copia silvestre de la proteína, pero también puede postularse la posibilidad de que sean mutaciones que puedan estar asociadas a la EP y no hallan sido descritas todavía como tal. Los estudios de asociación de los polimorfismos de LRRK2 han mostrado la existencia de diferencias importantes en la penetrancia de la presencia y el efecto de las mutaciones patogénicas con relación a la EP y a sus síntomas, según la raza e incluso según la población estudiada (Kumari y Tan, 2009). Esto hace plantearse la posibilidad de que algunas mutaciones de efecto pronunciado en la estructura como N363S (armadillo), R793M (anquirina) y R1067Q (LRR) puedan estar relacionadas con la patogénesis de parkinsonismos en alguna población.

Por último, destacar que muchas de las mutaciones aquí analizadas podrían afectar a la unión o a la posición de los distintos dominios entre sí. No obstante, el tamaño y la complejidad de la proteína no nos ha permitido realizar un análisis *in silico* de la proteína completa ni tan siquiera de la combinación de dos o tres dominios, con buenos resultados. Además, no debemos olvidar que los resultados obtenidos se basan en predicciones que toman como modelo otras proteínas ya descritas, por lo que la determinación de la estructura real de LRRK2 sigue siendo un objetivo para la elucidación de su implicación en la patogénesis de la EP.

Las figuras de aquellos polimorfismos y mutaciones patogénicas para las que no se han encontrado cambios a nivel estructural se muestran en el Anexo 6.

### **3.3. Clonación de dominios de LRRK2**

Debido a los resultados del análisis *in silico* realizado respecto a los límites de los dominios de LRRK2, se han re-clonado los dominios COR y ROC-COR, con el objetivo de repetir los ensayos de solubilidad y estabilidad proteica con estas nuevas definiciones

de dominios mejor acotadas. El procedimiento seguido para este fin se describe en el apartado 2.2 de Materiales y Métodos.

## 4. CONCLUSIONES

---

- 1- Los dominios LRR, ROCO, COR y KIN, tal como han sido definidos inicialmente en este trabajo, y expresados en un sistema procariota utilizando el plásmido pPROEX-HTa, no se encuentran plegados en una estructura soluble. Los dominios replegados no presentan la estabilidad requerida para realizar estudios estructurales. Estos resultados hacen dudar de la correcta definición de los límites de los dominios o de la idoneidad del sistema de expresión utilizado. Una redefinición y reclonación y/o la expresión en eucariotas de dominios (o combinaciones de los mismos) se plantean como futuros experimentos, necesarios para el avance en el conocimiento de la estructura y la función de esta proteína.
  
- 2- El análisis *in silico* de la estructura de LRRK2 es una herramienta útil para predecir la estructura de la proteína, así como el efecto que los cambios de amino ácido descritos (polimorfismos o mutaciones) producen en la misma. La mayoría de las mutaciones implicadas en la EP, así como muchos de los polimorfismos descritos, afectan a la estructura de la proteína (superficie, puentes de hidrógeno con otros residuos o distribución de cargas) o bien a los sitios de fosforilación de la misma, implicados en la regulación de su actividad. Los resultados obtenidos mediante estos análisis sirven de punto de partida y de hipótesis de trabajo para futuros experimentos en el laboratorio.





## 5. BIBLIOGRAFÍA

---

- Biskup S, Gerlach M, Kupsch A, Reichmann H, Riederer P, Vieregge P, Wüllner U, Gasser T. Genes associated with Parkinson syndrome. *Journal of Neurology*, 255: 9-17; 2008.
- Biskup S, Moore DJ, Celsi F, Higahsi S, West AB, Andrabi SA, Kurkinen K, Yu SW, Savitt JM, Waldvogel HJ, Faull RLM, Emson PC, Torp R, Ottersen OP, Dawson TM, Dawson VL. Localization of LRRK2 to membranous and vesicular structures in mammalian brain. *Ann Neurol*, 60: 557-569; 2006.
- Cole C, Barber JD, Barton GJ. The Jpred 3 secondary structure prediction server. *Nucleic Acids Res*, 36 (web server issue): 197-201; 2008.
- Dächsel JC, Taylor JP, Mok SS, Ross OA, Hinkle KM, Bailey RM, Hines JH, Szutu J, Madden B, Petrucelli L, Farrer MJ. Identification of potential protein interactors of Lrrk2. *Parkinsonism and Related Disorders*, 13: 382-385; 2007.
- De Lau LM y Breteler MM. Epidemiology of Parkinson's disease. *Lancet Neurol*, 5: 525-35; 2006.
- De Rijk MC, Launer LJ, Berger K, Breteler MM, Dartigues JF, Baldereschi M, Fratiglioni L, Lobo A, Martinez-Lage J, Trenkwalderand C. Hofman. Prevalence of Parkinson's disease in Europe: A collaborative study of population-based cohorts. Neurologic Diseases in the Elderly Research Group. *Neurology*, 54: S21-3; 2000.
- Deng J, Lewis PA, Greggio E, Sluch E, Beilina A, Cookson MR. Structure of the ROC domain from the Parkinson's disease-associated leucine-rich repeat kinase 2 reveals a dimeric GTPase. *Proc Natl Acad Sci U S A*, 105(5): 1499-504; 2008.
- Engebrecht J, Brent R, Kaderbhai MA. Minipreps of Plasmid DNA. *Current Protocols in Molecular Biology*, 1.6.1-1.6.10; 1991.

- Forneris F, Orru R, Bonivento D, Chiarelli LR, Mattevi A. ThermoFAD, a ThermoFluor-adapted flavin ad hoc detection system for protein folding and ligand binding. *FEBS J*, 276 (10): 2833-40; 2009.
- Gallagher SR. One-Dimensional SDS Gel Electrophoresis of Proteins. *Current Protocols in Protein Science*, 10.1.1-10.1.34; 1995.
- Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD, Bairoch A. Protein Identification and Analysis Tools on the ExPASy Server.
- The Proteomics Protocols Handbook, Humana Press, 571-607; 2005.
- Golbe LI. The genetics of Parkinson's disease: a reconsideration. *Neurology*, 40: S7-14; 1990.
- Gotthardt K, Weyand M, Kortholt A, Van Haastert PJ, Wittinghofer A. Structure of the Roc-COR domain tandem of *C. tepidum*, a prokaryotic homologue of the human LRRK2 Parkinson kinase. *EMBO J*, 27(16): 2239-49; 2008.
- Greggio E, Cookson MR. Leucine-rich repeat kinase 2 mutations and Parkinson's disease: three questions. *ASN Neuro*; 2009.
- Greggio E, Jain S, Kingsbury A, Bandopadhyay R, Lewis P, Kaganovich A, van der Brug MP, Beilina A, Blackinton J, Thomas KJ, Ahmad R, Miller DW, Kesavapany S, Singleton A, Lees A, Harvey RJ, Harvey K, Cookson MR. Kinase activity is required for the toxic effects of mutant LRRK2/dardarin. *Neurobiology of disease*, 23: 329-341; 2006.
- Greggio E, Zambrano I, Kaganovich A, Beilina A, Taymans JM, Daniëls V, Lewis P, Jain S, Ding J, Syed A, Thomas KJ, Baekelandt V, Cookson MR. The Parkinson disease-associated leucine-rich repeat kinase 2 (LRRK2) is a dimer that undergoes intramolecular autophosphorylation. *J Biol Chem*, 283(24):16906-14; 2008.
- Guo L, Gandhi PN, Wang W, Petersen RB, Wilson-Delfosse AL, Chen SG. The Parkinson's disease-associated protein, leucine-rich repeat kinase 2 (LRRK2), is an

authentic GTPase that stimulates kinase activity. *Exp Cell Res*, 313(16):3658-70; 2007.

- Häbig K, Walter M, Poths S, Riess O, Bonin M. RNA interference of LRRK2-microarray expression analysis of a Parkinson's disease key player. *Neurogenetics*, 9: 83-94; 2008.
- Hatano T, Kubo S, Imai S, Maeda M, Ishikawa K, Mizumo Y, Hattori N. Leucine-rich repeat kinase 2 associates with lipid rafts. *Hum Mol Genet*, 16: 678-690; 2007.
- Huang X, Chenand PC, Poole C. APOE-[epsilon]2 allele associated with higher prevalence of sporadic Parkinson disease. *Neurology*, 62: 2198-2002; 2004.
- Ito G, Okai T, Fujino G, Takeda K, Ichijo H, Katada T, Iwatsubo T. GTP binding is essential to the protein kinase activity of LRRK2, a causative gene product for familial Parkinson's disease. *Biochemistry*, 46:1380-1388; 2007.
- Jaleel M, Nichols RJ, Deak M, Campbell DG, Gillardon F, Knebel A, Alessi DR. LRRK2 phosphorylates moesin at threonine-558 characterization af how Parkinson's disease mutants affect kinase activity. *Biochem. J.* 405: 307-317; 2007.
- Jang H, Boltz DA, Webster RG, Smeyne RJ. Viral parkinsonism. *Biochimica et biophysica acta*, 1792(7): 714-21; 2008.
- Khandhar SM, Marks WJ. Epidemiology of Parkinson's disease. *Disease-a-month*, 53(4): 200-5; 2007.
- Klein CL, Rovelli G, Springer W, Schall C, Gasser T, Kahle PJ. Homo- and hetrodimerization of ROCO kinases: LRRK2 kinase inhibition by the LRRK2 ROCO fragment. *Journal of Neurochemistry*; 2009 (Epub ahead of print).
- Korr D, Toschi L, Donner P, Pohlez HD, Kreft B, Weiss B. LRRK1 protein kinase activity is stimulated upon binding of GTP to its Roc domain. *Cellular Signalling*, 18: 910-920; 2006.

- Kumari U, Tan EK. LRRK2 in Parkinson's disease: genetic and clinical studies from patients. *The FEBS Journal*, 276: 6455-6463; 2009.
- Lee SB, Kim W, Lee S, Chung J. Loss of LRRK2/PARK8 induces degeneration of dopaminergic neurons in *Drosophila*. *Biochemical and Biophysical Research Communications*, 358: 534-539; 2007.
- Lewitt PA. Levodopa for the treatment of Parkinson's disease. *The New England journal of medicine*, 359(23): 2468-76; 2008.
- Luzón Toro B, Rubio de la Torre E, Delgado A, Pérez Tur J, Hilfiker S. Mechanistic insight into the dominant mode of Parkinson's disease-associated G2019S LRRK2 mutation. *Human Molecular Genetics*, 16: 2031-2039; 2007.
- Marín I, van Egmond WN, van Haastert PJ. The Roco protein family: a functional perspective. *FASEB J*, 22(9):3103-10; 2008.
- Marín I. Ancient origin of the Parkinson disease gene LRRK2. *J Mol Evol*, 67: 41-50; 2008.
- Mata IF, Wedemeyer WJ, Farrer MJ, Taylor JP, Gallo KA. LRRK2 in Parkinson's disease: protein domains and functional insights. *TRENDS in Neurosciences*, 29: 286-293; 2006.
- Melrose H. Update on the functional biology of Lrrk2. *Future Neurol*, 3 (6): 669-681; 2008.
- Olson BJSC, Markwell J. Assays for Determination of Protein Concentration. *Current Protocols in Protein Science*, 3.4.1-3.4.29; 2007.
- Paisan-Ruiz C, Jain S, Evans EW, Gilks WP, Simon J, Van Der Brug M, Lopez De Munain A, Aparicio S, Gil AM, Khan N, Johnson J, Martinez JR, Nicholl D, Carrera IM, Pena AS, De Silva R, Lees A, Marti-Masso JF, Perez-Tur J, Woodand NW, Singleton AB. Cloning of the gene containing mutations that cause PARK8- linked Parkinson's disease. *Neuron*, 44: 595-600; 2004.

- Saha S, Guillily MD, Ferree A, Lanceta J, Chan D, Ghosh J, Hsu CH, Segal L, Raghavan K, Matsumoto K, Hisamoto N, Kuwahara T, Iwatsubo T, Moore L, Goldstein L, Cookson M, Wolozin B. LRRK2 modulates vulnerability to mitochondrial dysfunction in *Caenorhabditis elegans*. *J Neurosci*, 29:9210-8; 2009.
- Sämman J, Hegermann J, von Gromoff E, Eimer S, Baumeister R, Schmidt E. *Caenorhabditis elegans* LRRK-1 and PINK-1 Act Antagonistically in Stress Response and Neurite Outgrowth. *J Biol Chem*, 284: 16482-16491; 2009.
- Santpere G, Ferrer I. LRRK2 and neurodegeneration. *Acta Neuropathologica*, 117: 227-246; 2009.
- Schwede T, Kopp J, Guex N, Peitsch MC. SWISS-MODEL: An automated protein homology-modeling server. *Nucleic Acid Res*, 31 (13): 3381-5; 2003.
- Sen S, Webber PJ, West AB. Leucine-Rich Repeat Kinase 2 (LRRK2) Kinase Activity: Dependence on Dimerization. *J Biomol Chem*; 2009 (Epub ahead of print).
- Shin N, Jeong H, Kwon J, Heo HY, Kwon JJ, Yun HJ, Kim C, Han BS, Tong Y, She J, Hatano T, Hattori N, Kim K, Cahng S, Seol W. LRRK2 regulates synaptic vesicle endocytosis. *Experimental cell research*, 314(10):2055-65; 2008.
- Simón-Sánchez J, Herranz-Pérez V, Olucha-Bordonau F, Pérez-Tur J. LRRK2 is expressed in areas affected by Parkinson's disease in the adult mouse brain. *Eur J Neurosci*, 23 (3): 659-66; 2006.
- Smith WW, Pei Z, Jiang H, Moore DJ, Liang Y, West AB, Dawson VL, Dawson TM, Ross CA. Leucine-rich repeat kinase 2 (LRRK2) interacts with parkin, and mutant LRRK2 induces neuronal degeneration. *PNAS*, 102: 18676-18681; 2005.
- Taylor JP, Hulihan MM, Kachergus JM, Melrose HL, Lincoln SJ, Hinkle KM, Stone JT, Ross OA, Hauser R, Aasly J, Gasser T, Payami H, Wszolek ZK, Farrer MJ. Leucine – rich repeat kinase 1: a paralog of LRRK2 and a candidate gene for Parkinson's disease. *Neurogenetics*, 8: 95-102; 2007.

- Tomiyama H, Mizuta I, Li Y, Funayama M, Yoshino H, Li L, Murata M, Yamamoto M, Kubo S, Mizuno Y, Toda T, Hattori N. LRRK2 P755L variant in sporadic Parkinson's disease. *J Hum Genet*, 53(11-12): 1012-5; 2008.
- Tong Y, Pisani A, Martella G, Karounai M, Yamaguchi H, Pothos EN, Shen J. R1441C mutation in LRRK2 impairs dopaminergic neurotransmission in mice. *PNAS*, 2009.
- van Egmond WN, Kortholt A, Plak K, Bosgraaf L, Bosgraaf S, Keizer-Gunnink I, van Haastert PJ. Intramolecular activation mechanism of the Dictyostelium LRRK2 homolog Roco protein GbpC. *J Biol Chem*, 283(44): 30412-20; 2008.
- Voytas D. Agarose Gel Electrophoresis. *Current Protocols in Molecular Biology*, 2.5A.1-2.5A.9; 2000.
- Wakabayashi K, Tanji K, Moriand F y Takahashi H. The Lewy body in Parkinson's disease: molecules implicated in the formation and degradation of alpha-synuclein aggregates. *Neuropathology*, 27: 494-506. 2007.
- Waters CH, Miller CA. Autosomal dominant Lewy body parkinsonism in a four-generation family. *Annals of Neurology*, 35(1): 59-64; 1994.
- Webber PJ, West AB. LRRK2 in Parkinson's disease: function in cells and neurodegeneration. *FEBS J*, 2009 (Epub ahead of print).
- West AB, Moore DJ, Biskup A, Bugayenko A, Smith WW, Ross CA, Dawson VL, Dawson TM. Parkinson's disease-associated mutations in leucine-rich repeat kinase 2 augment kinase activity. *PNAS*, 102: 16842-16847; 2005.
- West AB, Moore DJ, Biskup S, Bugayenko A, Smith WW, Ross CA, Dawson VL, Dawson TM. Parkinson's disease-associated mutations in leucine-rich repeat kinase 2 augment kinase activity. *Proc Natl Acad Sci U S A.*, 102 (46): 16842-7; 2005.
- West AB, Moore DJ, Choi C, Andrabi SA, Li X, Dikeman D, Biskup S, Zhang Z, Lim K, Dawson VL, Dawson TM. Parkinson's disease-associated mutations in LRRK2 link

enhanced GTP-binding and kinase activities to neuronal toxicity. *Human Molecular Genetics*, 16: 223-232; 2007.

- Wu S, Zhang Y. LOMETS: A local meta-threading-server for protein structure prediction. *Nucleic Acids Research*, 35: 3375-3382; 2007.
- Yang YX, Wood NW, Latchman DS. Molecular basis of Parkinson's disease. *Neuroreport*, 20: 150-156; 2009.
- Yue Z. LRRK2 in Parkinson's disease: in vivo models and approaches for understanding pathogenic roles. *FEBS J*; 2009.





## **ANEXOS**

---



## Anexo 1\_ Cuantificación proteica

Tabla A1.1. Cuantificación mediante  $Abs_{280}$  de proteína desnaturalizada según la concentración de urea.

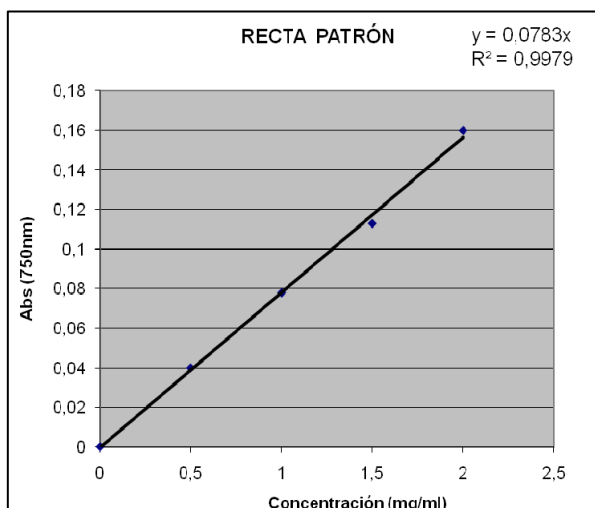
Tampón urea	Construcción	$Abs_{280}$	$(Abs_{280}/\epsilon) \cdot dil \cdot Mr$	Concentración
8M	COR	0,95	$(0,95/37.025) \cdot 100 \cdot 27.687$	71 mg/mL
	ROCO	0,93	$(0,93/59.485) \cdot 100 \cdot 47.813,8$	74,8 mg/mL
6M	COR	0,92	$(0,92/37.025) \cdot 100 \cdot 27.687$	68,8 mg/mL
	ROCO	0,92	$(0,92/59.485) \cdot 100 \cdot 47.813,8$	74 mg/mL
4M	COR	0,93	$(0,93/37.025) \cdot 20 \cdot 27.687$	14 mg/mL
	ROCO	No se considera soluble en esta concentración de urea		

Se calcula la concentración mediante la ecuación de Lamber-Beer ( $Abs = l \cdot \epsilon \cdot c$  siendo  $l = 1$  cm) teniendo en cuenta el factor de dilución de la muestra. Cabe destacar que los valores de coeficientes de extinción ( $\epsilon$ ) utilizados son predichos, por lo que estos resultados sirven, más que para cuantificar de forma rigurosa, para comparar la concentración de proteína en los diferentes tampones.

Tabla/Figura A1.2. Cuantificación mediante Bradford de COR soluble tras su replegado en el tampón 13. (A) Cálculo de la recta patrón utilizando un estándar de BSA de concentración conocida (B) Cálculo de la concentración de la muestra problema por la extrapolación con la recta patrón anterior.

(A)

RECTA PATRÓN (con BSA)	
Abs (750 nm)	Concentración (mg/mL)
0	0
0,04	0,5
0,078	1
0,113	1,5
0,16	2



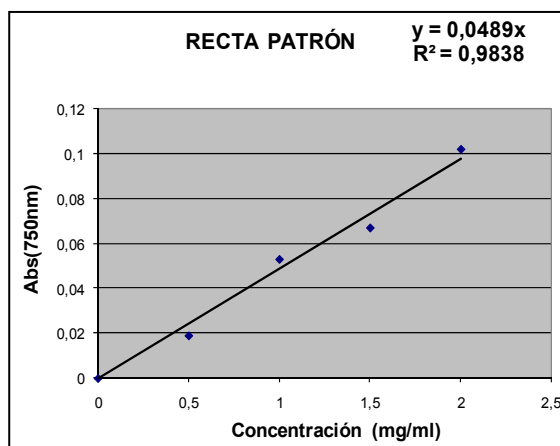
(B)

MUESTRA		
Abs (750 nm)	Dilución	Concentración (mg/mL)
0,17	1/10	<b>21,63</b>
0,082	1/20	<b>20,87</b>

**Tabla/Figura A1.3. Cuantificación del extracto de cerebro de rata para el *pull-down* mediante Bradford.** (A) Cálculo de la recta patrón utilizando un estándar de BSA de concentración conocida (B) Cálculo de la concentración de la muestra problema por la extrapolación con la recta patrón anterior.

(A)

RECTA PATRÓN (con BSA)	
Abs (750 nm)	Concentración (mg/mL)
0	0
0,019	0,5
0,053	1
0,067	1,5
0,102	2



(B)

MUESTRA		
Abs (750 nm)	Dilución	Concentración (mg/mL)
0,07	1/100	<b>142,2</b>
0,034	1/200	<b>138,2</b>

## Anexo 2\_ Calibrado de la cromatografía de exclusión molecular

Tabla A2.1. Calibrado de la columna de exclusión molecular S-200.

	vo (mL)	ve (mL)	ve-vo	Mm (kDa)	LnMm
azul dextrano	8,35	8,35	0	2000	7,6
b-amilasa	8,35	12,15	3,8	200	5,3
citrocromo c	8,35	18,21	9,86	12,4	2,52
alcohol deshidrogenasa	8,35	13,08	4,73	150	5,01
anhidrasa carbónica	8,35	16,6	8,25	29	3,37
albúmina	8,35	14,12	5,77	66	4,19

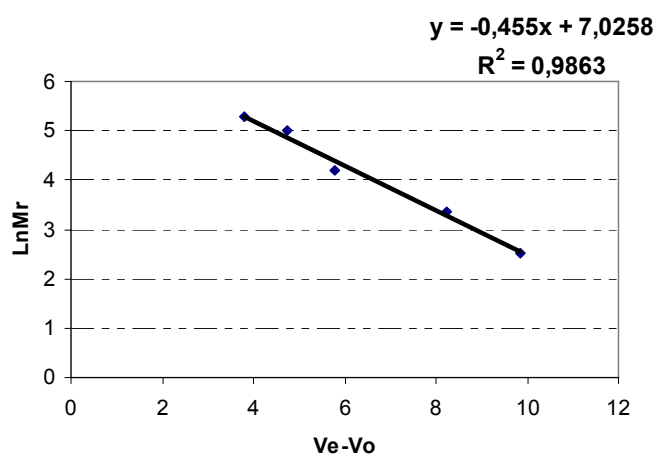


Figura A2.1. Recta patrón de la cromatografía de exclusión molecular.

ve: volumen de exclusión vo: volumen inicial (sustancia de mayor Mm). La diferencia  $ve - vo$  y el logaritmo neperiano (Ln) de la Mm para cada compuesto presenta una relación lineal. Con la recta patrón obtenida se calcula la Mm de la solución proteica de COR una vez replegado tal y como se muestra en la siguiente tabla.

Tabla A2.2. Cálculo de la Mm de la muestra de COR tras su replegado *in vitro* con el tampón 13.

ve (mL)	vo (mL)	ve-vo	Ln Mm	Mm (kDa)
11,8	8,35	3,45	5,46	235

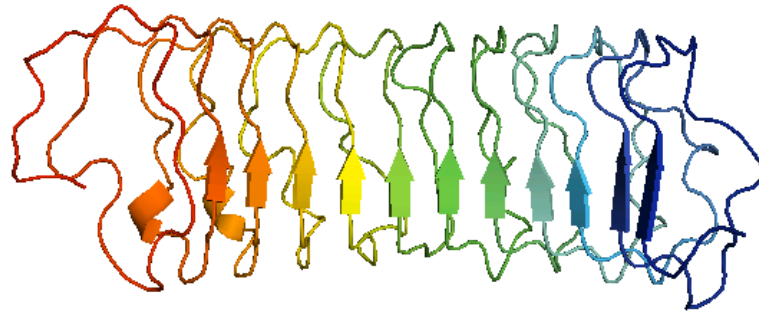


### Anexo 3\_ Tampones utilizados en la prueba de estabilidad proteica

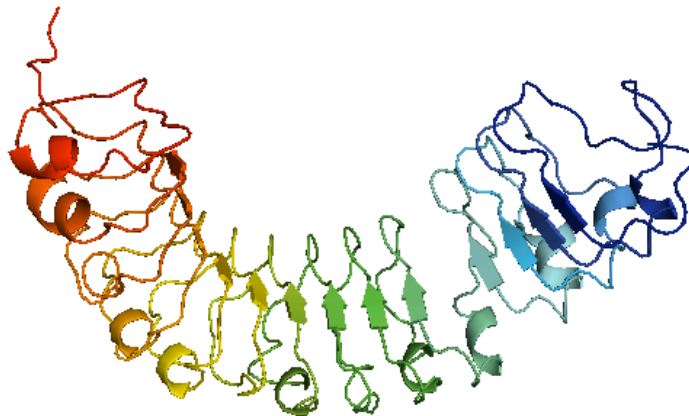
Tabla A3.1. Batería de tampones ensayados en la prueba de estabilidad proteica. Se muestra su posición en la placa de 96 pocillos, las posiciones que faltan son los duplicados.

	1	2	3	4	5	6	7	8	9	10	11	12
A	100 mM NaAc pH 5 150 mM NaCl	100 mM Pyridine pH 5,5 150 mM NaCl	100 mM Cacodilato pH 6 150 mM NaCl	100 mM MES pH 6 150 mM NaCl	100 mM NaCl pH 6,5 150 mM NaCl	100 mM Imidazole pH 7 50 mM NaCl	100 mM PIPES pH 7 150 mM NaCl	100 mM MOPS pH 7,2 50 mM NaCl	100 mM HEPES pH 7,5 150 mM NaCl	100 mM Tris pH 8 150 mM NaCl	100 mM Tricine pH 8 150 mM NaCl	100 mM Bicine pH 8,3 150 mM NaCl
B	100 mM Glicina/Na OH pH 9 150 mM NaCl	65 mM Bicine pH 9 150 mM NaCl	100 mM CHES pH 9,5 150 mM NaCl	100 mM Ebscon pH 9,5 150 mM NaCl	100 mM CAPS pH 10 150 mM NaCl	100 mM HEPES pH 7,5 50 mM NaCl	100 mM HEPES pH 7,5 50 mM NaCl	100 mM HEPES pH 7,5 500 mM NaCl	100 mM HEPES pH 7,5 150 mM NaCl 10% Glicerol	100 mM HEPES pH 7,5 150 mM NaCl 10% Glicerol	100 mM HEPES pH 7,5 500 mM NaCl 5% Glicerol	100 mM HEPES pH 7,5 150 mM NaCl 50 mM L- Arg-50 mM L-glu
C	100 mM HEPES pH 7,5 150 mM NaCl 50 mM L- Arg	100 mM HEPES pH 7,5 150 mM NaCl 100 mM LUC	100 mM HEPES pH 7,5 150 mM KCl	100 mM HEPES pH 7,5 150 mM NaCl	100 mM HEPES pH 7,5 150 mM NaCl 0,1 mM NAC	100 mM HEPES pH 7,5 150 mM NaCl 5 mM MgCl2	100 mM HEPES pH 7,5 150 mM NaCl 0,1 mM ZnCl2	100 mM HEPES pH 7,5 150 mM NaCl 0,1 mM CaCl2	100 mM HEPES pH 7,5 150 mM NaCl 0,1 mM CaCl2 + 0,1 mM CoCl2	100 mM HEPES pH 7,5 150 mM NaCl 0,1 mM MnCl2	100 mM HEPES pH 7,5 50 mM Na-K fosfato	100 mM HEPES pH 7,5 50 mM NH4Ac
D	100 mM HEPES pH 7,5 50 mM NH4SO4	100 mM HEPES pH 7,5 150 mM NaCl 50 mM Betaina	100 mM HEPES pH 7,5 150 mM NaCl 200 mM betaina	100 mM HEPES pH 7,5 150 mM NaCl 500 mM Sacarosa	100 mM HEPES pH 7,5 150 mM NaCl 100 mM Sacarosa	100 mM Tris, fosfato pH 7,5 150 mM NaCl	100 mM Tris pH 8 150 mM NaCl 5% Glicerol	100 mM Tris pH 8 150 mM NaCl 10% Glicerol	100 mM Tris pH 8 500 mM NaCl 5% Glicerol	100 mM MES pH 6 150 mM NaCl 5% Glicerol	100 mM MES pH 6 150 mM NaCl 10% Glicerol	100 mM MES pH 6 500 mM NaCl 5% Glicerol

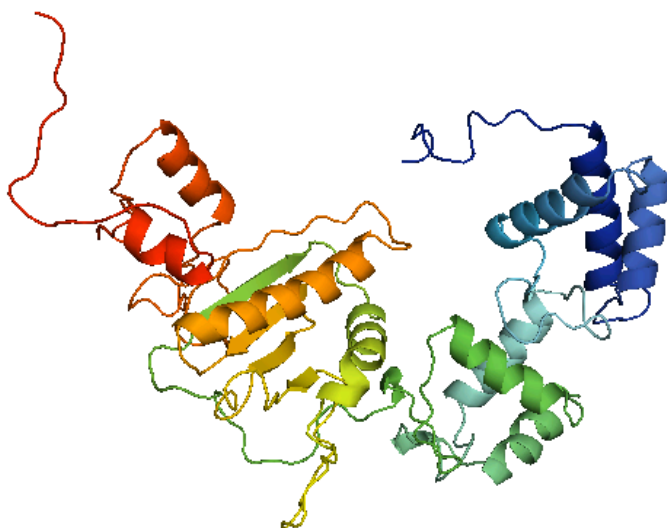
## Anexo 4 Análisis in silico: definición de dominios



**Figura A4.1. Representación de la estructura terciaria del dominio LRR (aas 1037-1313).** Se observa que solo posee 11 láminas- $\beta$ , cuando se ha descrito que este dominio tiene 13, además sus extremos están desestructurados.

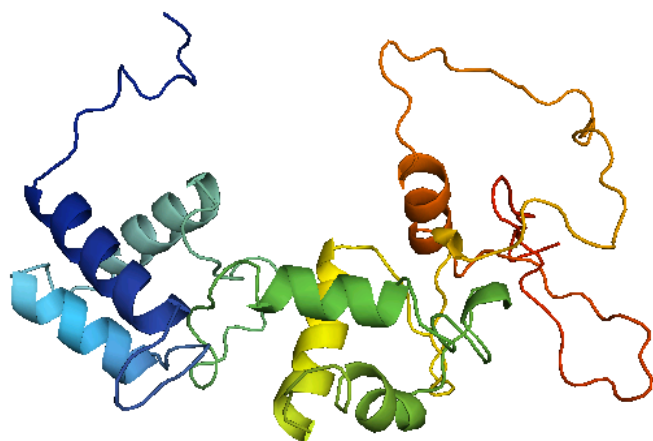


**Figura A4.2. Representación de la estructura terciaria del dominio LRR (aas 979-1310).** Se observa que no posee una estructura curva correcta y que, además sus extremos están desestructurados.

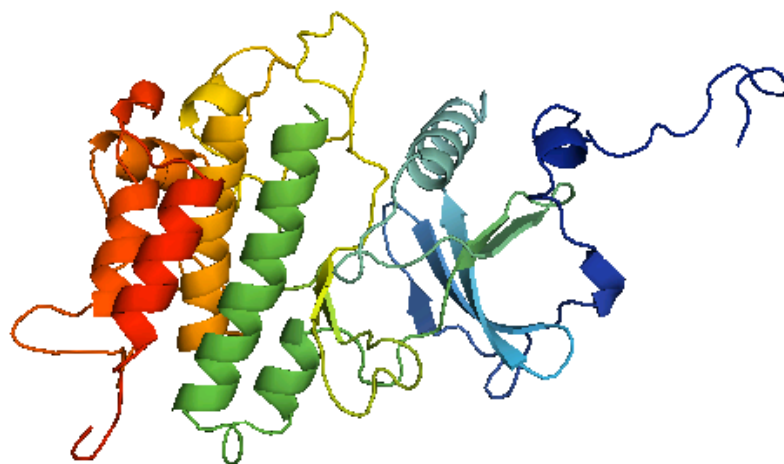


**Figura A4.3. Representación de la estructura terciaria del dominio COR (aas 1511-1878).** Aparece desestructurado a lo largo de toda su estructura y principalmente en los extremos.





**Figura A4.4. Representación de la estructura terciaria del dominio COR (aas 1513-1745).** Aparece desestructurado a lo largo de toda su estructura y principalmente en los extremos.



**Figura A4.5. Representación de la estructura terciaria del dominio quinasa (aas 1845-2143).** Los extremos del dominio así como el sitio catalítico y el lazo de activación aparecen desestructurados.



**Figura A4.6. Representación de la estructura terciaria del dominio WD40 (aas 2168-2387).** Sólo presenta cinco de las siete repeticiones características de este tipo de dominio.

## **Anexo 5\_ Análisis *in silico*: estructura secundaria (JPred3)**

Leyenda: “H”: hélice- $\alpha$ , “E”: lámina- $\beta$ , “-”: lazo. “B” (buried): posiciones menos accesibles al solvente. Fiabilidad de la predicción: valores del 0 al 9. Las predicciones de menos de 5 posiciones no se consideraran fiables.

**ARMADILLO** (aas 150-510)

```
Jnet      :--EEE-----HHHHHHHHHHHH--HHHHHHHHHHHHHHHHHH-----HHHHHHHH--HHHHHHHHHHHH--
HHHHHHHHHHHHHHHHHH-----HHHHHHH--HHHHHHHHHHH--HHHHHHHHHHHHHHHHHH-----HHHHHH--
HHHHHHHHHHHHH--HHHHHHHHHHHHHHHHHHHHHH-----HHH-----HHHH--HHHHHHHHHHHH--
HHHHHHHHHHHHHHHHHHHH-----HHHHHHHH--HHHHHHHHHHHHHH--HHHHHHHHHHHHHHHHHHHH--HHHHHHHHH--
HHHHHHHHHHHH--HHHHHHHHHHHHHHHHHHHH--HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH--HHHHHHHHHHHH-- : Jnet
```

```

Jnet_5      : B-----BB-BB-BB-----BB-BB-BB-BB-----B--B--BB-BB-----BB-BB-BBBBBB-----
-B-----BB-BB-BB-----BB-BB-BB-BB-----B--B--B--BB-----BB-BBBBBB--BB-----B-----BB-
BB-BB-----BB-BB-BBBB-BB-----B--BB-B-----BB-BB-BBBBBB-----B--B--BB-BB-----BB-
BB-BB-BBB-----B--BB-BB-B-----BB-BB-B-- : Jnet_5

```

```

Jnet Rel      :
000000087568899999745886689999999987408872017888745765899999974588668999998877630
068763677774478608899999800788668999999998860688615566247659999999745886689999999
9998700367770000067776550000104500999999985178658999999999886268885023554106870478
99999874488856899999999998744885689999885460799999986078866899999999998607860099999
999999999874588668999999998607 : Jnet Rel

```

**Figura A5.1. Predicción de la estructura secundaria del dominio armadillo de LRRK2.**

**ANKIRINA** (aas 690-860)

```
Jnet      : HHHHHHHHHHHH-----HHHHHHHHH---HHHHHHHHHH-----HHHHHHHH--
HHHHHHHHHH-----HHHHHHHH--HHHHHHHHHH-----HHHHHHHH--HHHHHHHHHH-----
HHHHHHHH--HHHHHHHHHH-- : Jnet
```

```

Jnet_5      : -B-BB-----BBBB-BB----BB-BB-----BBBBBB----BB-BB-B-----BBB-BB--
---BB-BB-----BBBBBB----BB-BBB-----BB-BB----BB-B-----: Jnet_5

```

Jnet Rel :  
05799999000023678775107899986166589998874267777777777535788884577689999874268888775  
3578888457768999987426777777777653456663078003545552367777777776357899861665899987  
50899 : Jnet Rel

**Figura A5.2. Predicción de la estructura secundaria del dominio anquirina de LRRK2.**

**LRRs (aas 984-1278)**

```

Jnet      : --EEE-----EEEEEE-----HHH-----EEEE-----
-----
           HHHH-----EE-----
-----
HHHHHHHHHHHHHHHHHHHH-----EEEE-----HHHH-----EEEE--- : Jnet

```



```

-----HHHHHHHHHHHHHHHH-----HHHHHHHHHHHH-----HHHHHHHHHHHHHHHH-----HHHHHHHHHH-----
: Jnet

Jnet_5      : -----B-B-B-----BBB-B-----B--BB-B-B---BB-B--BB-----BBBBBB---B---B-----B--
BB--BB-BB--B---BBB-B-B-BBB-----B-BBBBBBB-----BBBBB-BBBBBB-----BBBBBBBBBBBBBBBBB-----
-----BB-B-----B-B-BB-BB-----B-B--B-----: Jnet_5

Jnet Rel      :
99840564257548999985475489987356008999999999988037888037899985486178873577645777640
58887468999999999999998703770552078873680004788873088860054444577873577610787625656
40478886100899999999986146788777760688999870267777777700689999999987404567774688
98886146789 : Jnet Rel

```

**Figura A5.6. Predicción de la estructura secundaria del dominio quinasa de LRRK2.**

**WD40 (aas 2142-2498)**

```

Jnet      : --EEE---EEEEEEEE-----EEEEEE---EEEEEE-----EEEE---EEEEEEEE-----EEEEEE---EEEEEE-
---HHEEEE-----EEEEEE-----EEEEEE---EEEEEE-----EEEE-----EEEEEE-----EEEEEEEE---EEEEEE---
EEEEEE---EEEE---EEEEEE---EEEEEE---EEEEEE---EEEEEE-----HEEEEEEE---EEEEEE---
EEEEEE---EEE---EEEEEEEE-----EEEEEE-----EEEEEE---: Jnet

Jnet_5    : ---B---B-BBBB-----BBBBB---B-B-B-----B-BB-BB-----BBBBBB---B-B-B-----
-----BBBB-B-----BBBBBB-B-BBB-----B-B---B-BBB-----BBBBB---B-B-----B-----
---B-BB---BBBB---B-B-----B-----B-BBBBBB---BBBBB---BBBBB-B---B-----B-BBB-----
----BBBB-----B-B---: Jnet_5

Jnet Rel      :
9636037870478885378885078987158884089988537888002680056278998614777877358986088568
99844760000000378885377753465665126744899980461788853677707777735775268777605650056
65446016898555078860886078998358872675008887358999616458998508846898716751231033135
67777777777777762110000000068805897065089986157777000000777744888885068887257761
477777777777767752677538 : Jnet Rel

```

**Figura A5.7. Predicción de la estructura secundaria del dominio WD40 de LRRK2.**

## Anexo 6 \_ *Análisis in silico: estructura terciaria*

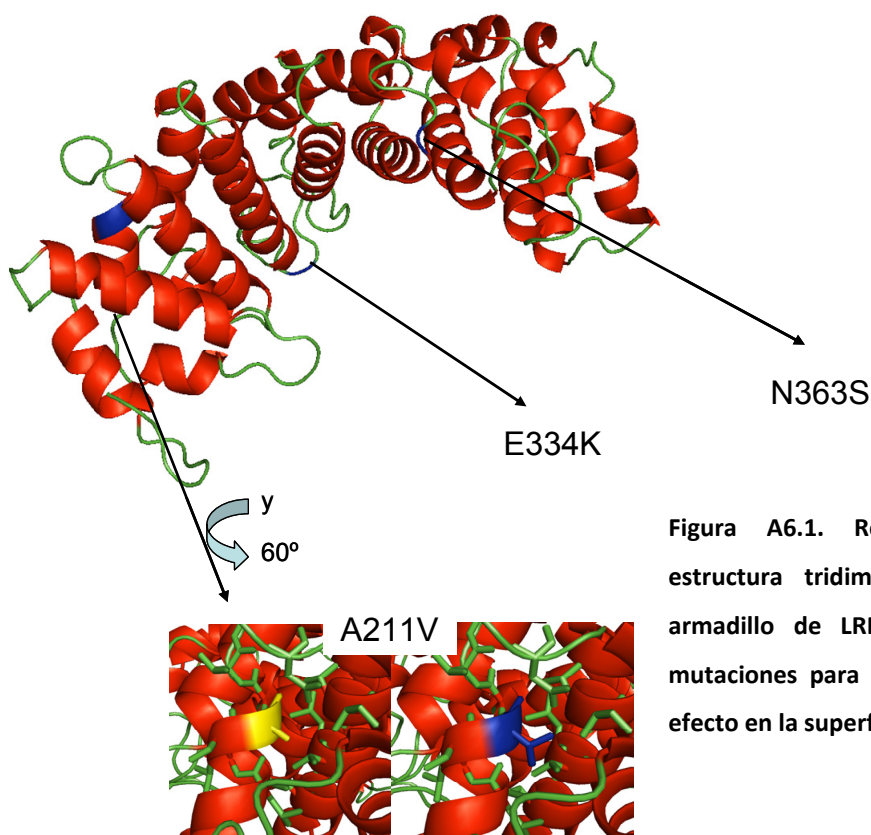


Figura A6.1. Representación de la estructura tridimensional del dominio armadillo de LRRK2 y detalle de las mutaciones para las que no se observa efecto en la superficie.

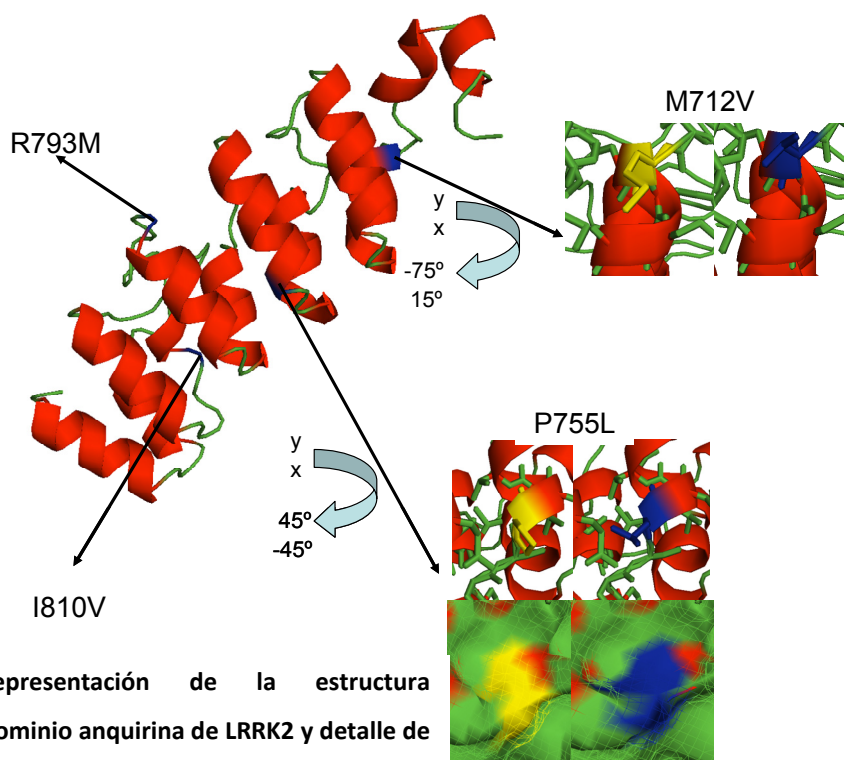


Figura A6.2. Representación de la estructura tridimensional del dominio anquirina de LRRK2 y detalle de las mutaciones para las que no se observa efecto en la superficie.

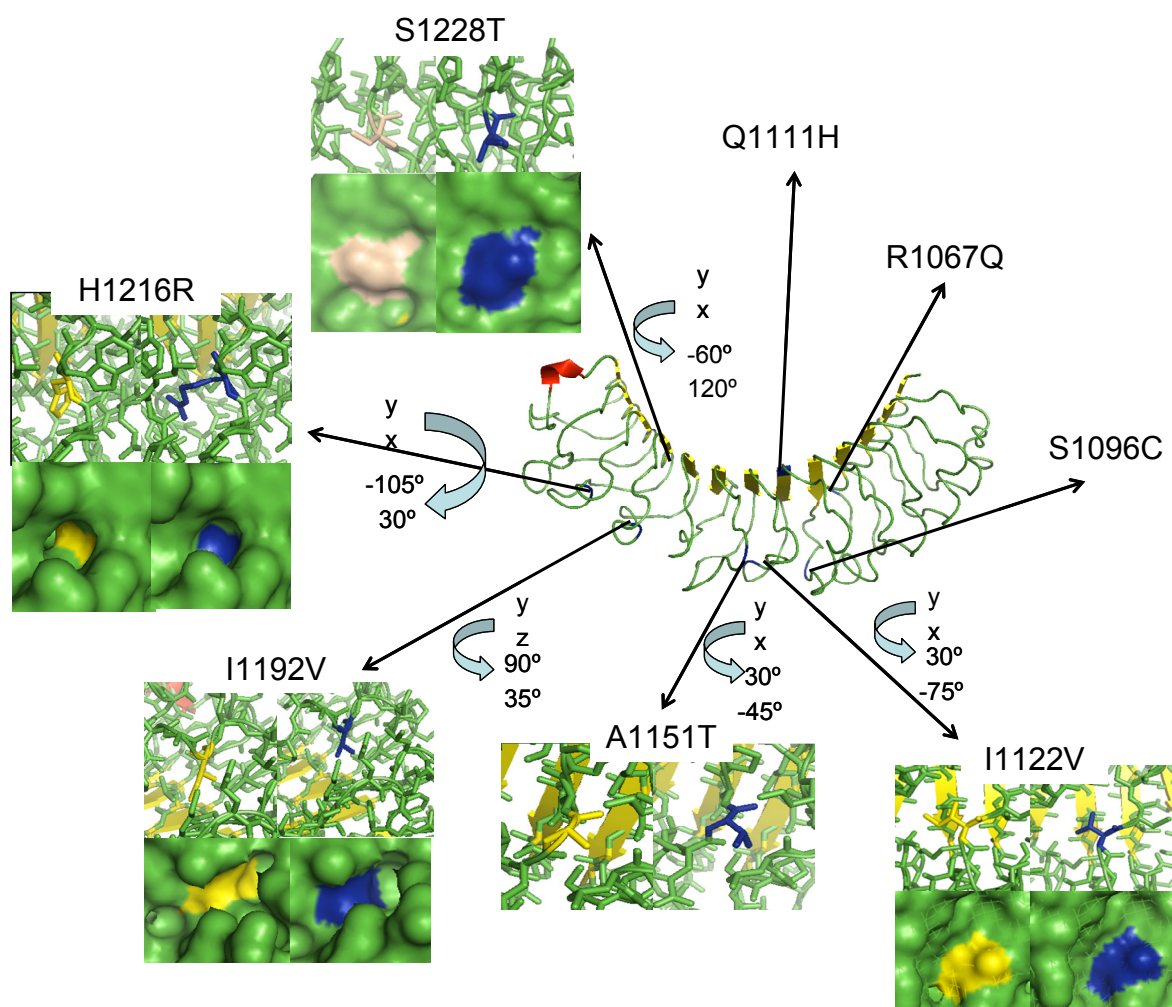


Figura A6.3. Representación de la estructura tridimensional del dominio LRR de LRRK2 y detalle de las mutaciones para las que no se observa efecto en la superficie.

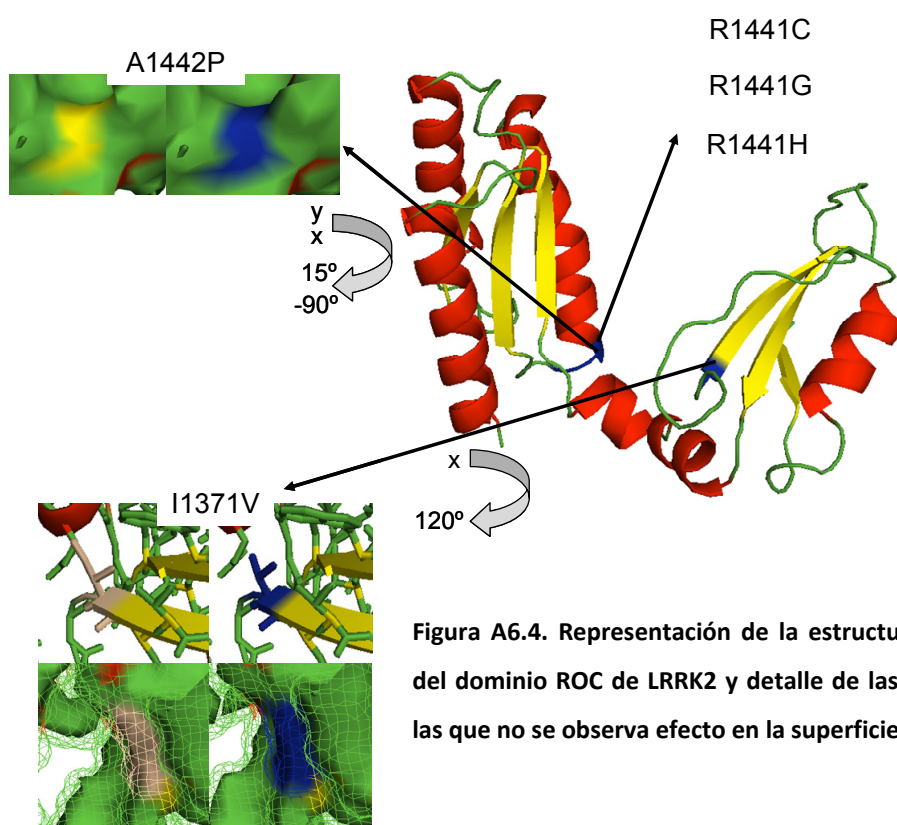


Figura A6.4. Representación de la estructura tridimensional del dominio ROC de LRRK2 y detalle de las mutaciones para las que no se observa efecto en la superficie.

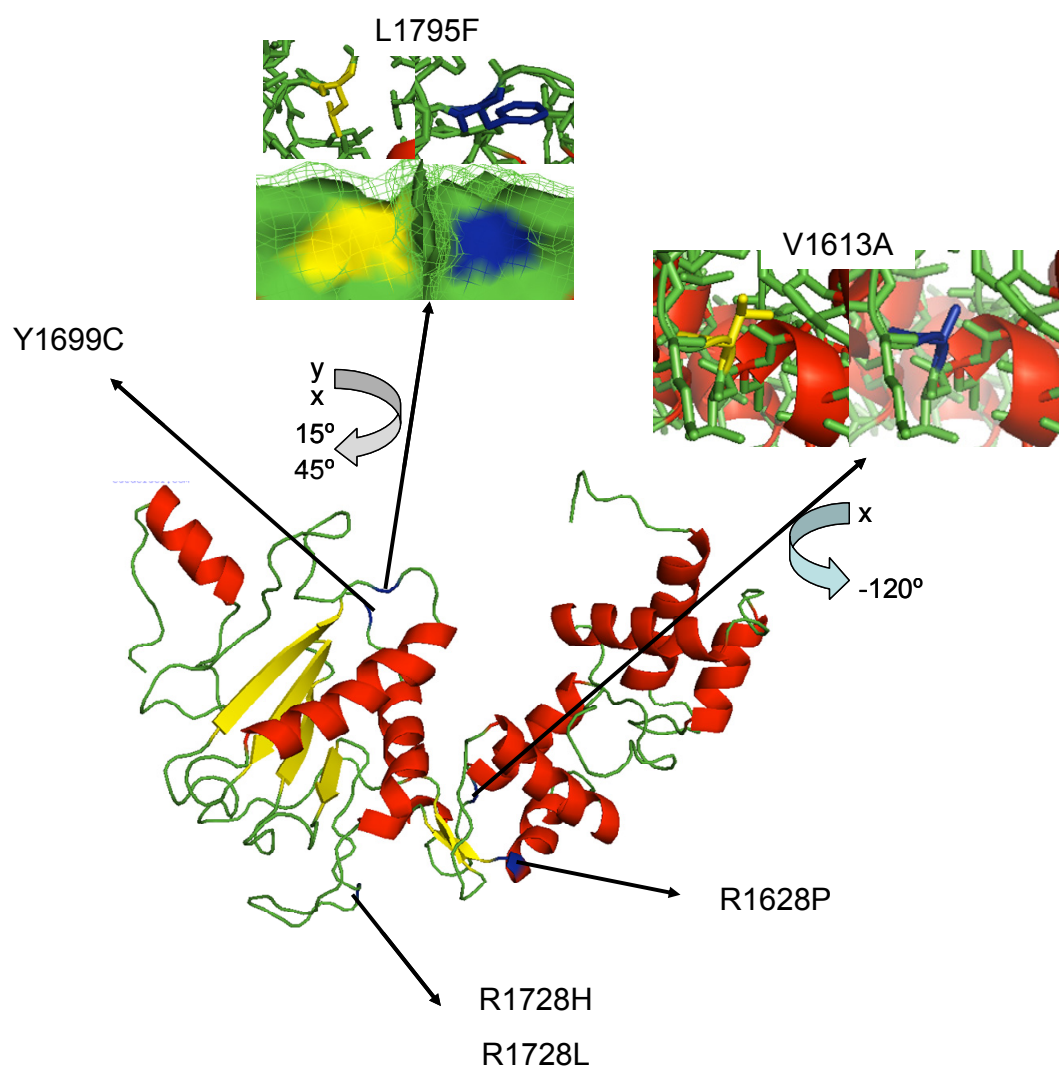


Figura A6.5. Representación de la estructura tridimensional del dominio COR de LRRK2 y detalle de las mutaciones para las que no se observa efecto en la superficie.



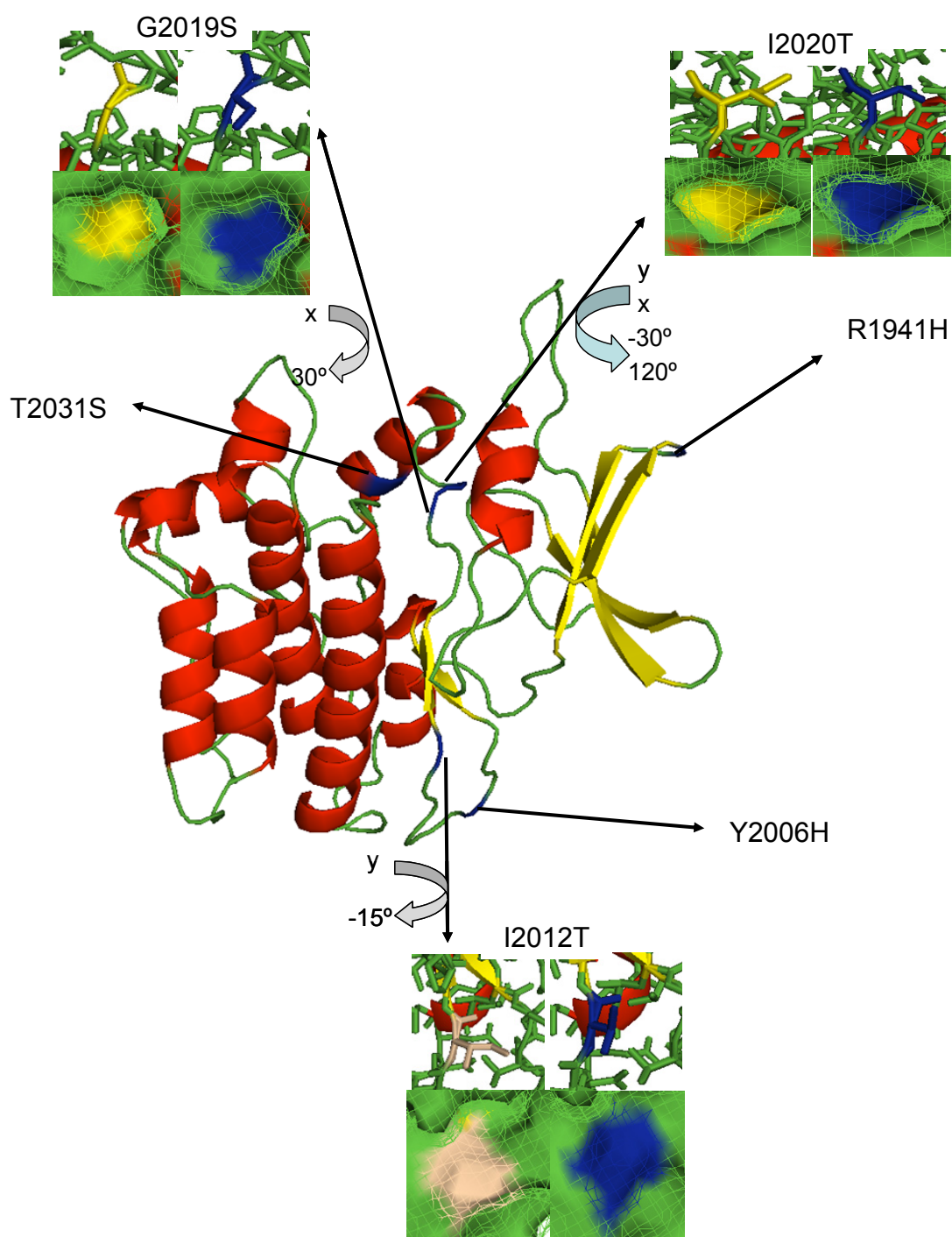
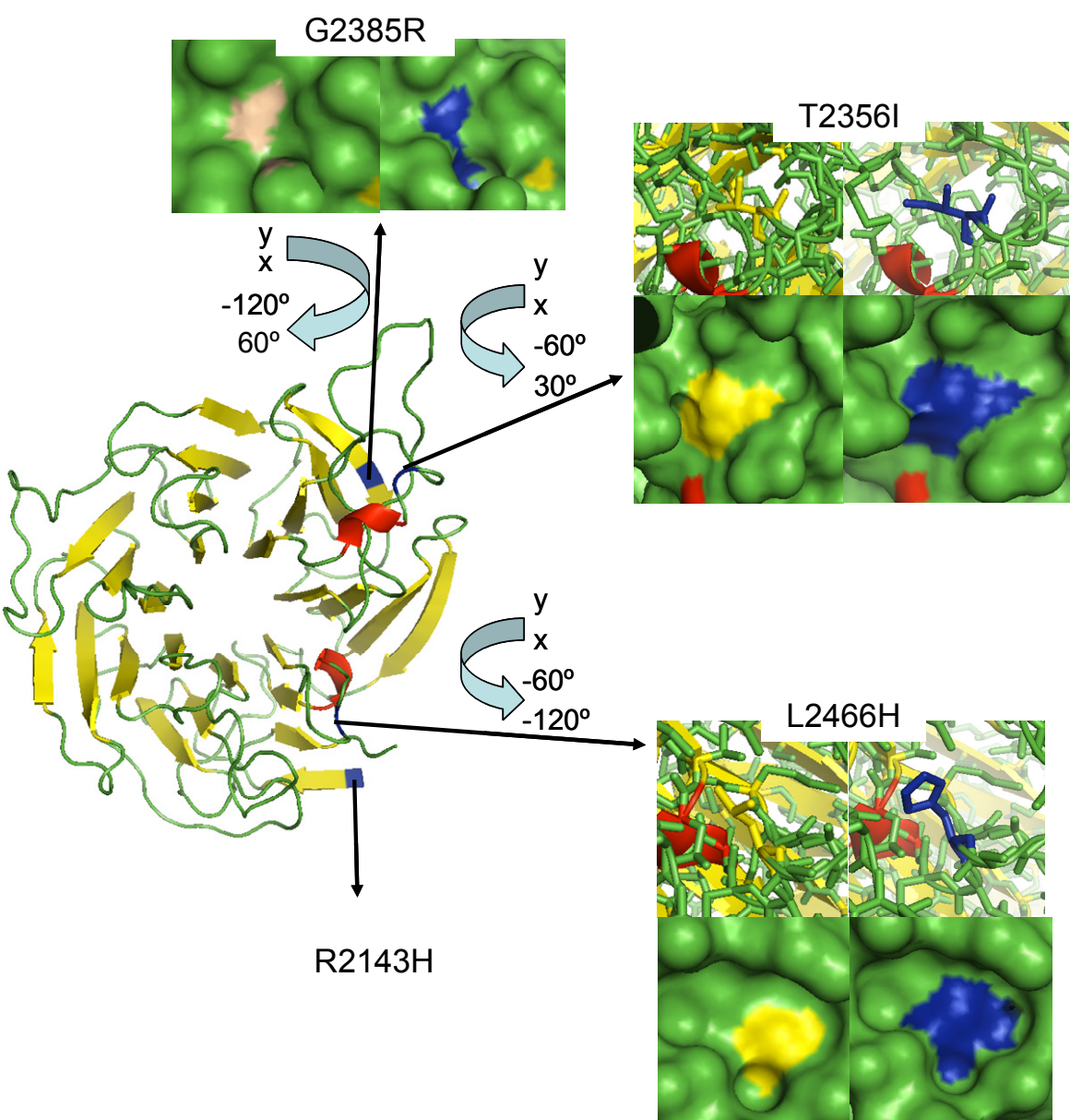


Figura A6.6. Representación de la estructura tridimensional del dominio quinasa de LRRK2 y detalle de las mutaciones para las que no se observa efecto en la superficie.





**Figura A6.7. Representación de la estructura tridimensional del dominio WD40 de LRRK2 y detalle de las mutaciones para las que no se observa efecto en la superficie.**